



UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INDUSTRIAS

**Evaluación Comparativa de Modelos GARCH y
GARCH-LSTM con VIX para la Predicción de
Volatilidad en ETFs del Sector Energético de EE. UU.**

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL**

AUTOR:

ALONSO MERCADO ROJEL

PROFESOR GUÍA:

Dr. WERNER KRISTJANPOLLER

PROFESOR CORREFERENTE:

Dr. BERNARDO PINCHEIRA

VALPARAÍSO, 10 DE DICIEMBRE DE 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo:

Evaluación comparativa de Modelos GARCH y GARCH-LSTM Incorporando el Índice VIX Para La Predicción de Volatilidad en ETFs del Sector Energético de EE.UU.

Nombre del candidato(a):

Alonso Mercado Rojas

Carrera / Grado:

Ingeniero Civil Industrial

Campus: Casa Central, Valparaíso Departamento: Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Werner Kristian Jeldes, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO** contiene información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por (marcar una opción):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria p tesis:

Fecha: 10/12/2025

Firma: [Firma]

Estudiante o Candidato(a):

Fecha: 10/12/2025

Firma: [Firma]

Resumen Ejecutivo

El presente estudio tiene como objetivo comparar el desempeño predictivo de modelos clásicos de volatilidad —pertenecientes a la familia GARCH— con modelos híbridos que integran redes neuronales LSTM (Long Short-Term Memory) aplicados a series de tiempo financieras de fondos cotizados en bolsa (ETFs) del sector energético de Estados Unidos. Se consideran nueve ETFs representativos, abarcando tanto energías renovables como convencionales, con el fin de analizar posibles diferencias en el comportamiento de la volatilidad entre ambos grupos.

La metodología combina un enfoque econométrico y de aprendizaje profundo. En una primera etapa, se estiman modelos GARCH, GJR-GARCH y EGARCH, con y sin inclusión del índice VIX como variable exógena. Posteriormente, se desarrollan modelos híbridos LSTM–GARCH, entrenados mediante optimización bayesiana con Optuna, considerando configuraciones de hiperparámetros que equilibran complejidad y estabilidad. La evaluación del desempeño se realiza bajo un esquema *out-of-sample* utilizando métricas como R^2 , MSE, MAE, QLIKE, HMSE, HMAE, MDA y el procedimiento de Model Confidence Set (MCS) de Hansen et al. (2011).

Los resultados evidencian que los modelos híbridos mejoran sistemáticamente las métricas de ajuste y error medio, particularmente en la predicción de la desviación estándar, donde alcanzan superioridad estadística según la prueba MCS. Sin embargo, en el modelado de la varianza, las diferencias con los modelos clásicos no resultan estadísticamente significativas, sugiriendo que la mayor complejidad de las arquitecturas LSTM no siempre se traduce en mejoras sustanciales. Asimismo, la inclusión del VIX como variable exógena aporta ligeras ventajas en estabilidad direccional (MDA), aunque no garantiza dominancia sobre los modelos sin dicha variable.

En cuanto al comportamiento por tipo de activo, los ETFs de energías renovables presentan una volatilidad más alta y heterogénea en el periodo de evaluación, mientras que los de energías convencionales muestran patrones más estables, sin que ello implique diferencias significativas en la capacidad predictiva de los modelos.

Índice

1. Problema de investigación	11
2. Objetivos de la investigación	12
2.1. Objetivo general	12
2.2. Objetivos específicos	12
3. Marco Teórico	13
3.1. Fundamentos del Mercado Bursátil	13
3.2. Mercado Bursátil de Estados Unidos	13
3.3. Fondos Cotizables en Bolsa (ETFs) Energéticos	14
3.4. Índice de Volatilidad (VIX)	16
3.5. Modelos de Varianza Condicional	16
3.5.1. Modelo ARCH	17
3.5.2. Modelo GARCH	17
3.5.3. Modelo EGARCH	18
3.5.4. Modelo GJR-GARCH	18
3.6. Modelos GARCH–VIX	19
3.7. Machine Learning	20
3.7.1. Redes Neuronales Recurrentes (RNN)	21
3.7.2. Redes LSTM	21
3.7.3. Modelos Híbridos LSTM–GARCH	22
3.7.4. Modelos Híbridos LSTM–GARCH–VIX	23
3.8. Optimizador Bayeseano	23
3.9. Métricas de Evaluación	24
4. Metodología	29
4.1. Análisis Exploratorio	29
4.2. División de Datos	32

4.3.	Entrenamiento de Modelos Clásicos	33
4.4.	Entrenamiento de Modelos Híbridos	33
4.4.1.	Estandarización de Variables	33
4.4.2.	Entrenamiento Supervisado	33
4.5.	Optimización de Modelos Clásicos	34
4.6.	Optimización de Modelos Híbridos	35
4.7.	Evaluación y Comparación	35
5.	Resultados	35
5.1.	Análisis Exploratorio	35
5.2.	FAN	39
5.2.1.	Entrenamiento modelos familia GARCH FAN (σ^2)	39
5.2.2.	Entrenamiento modelos Híbridos Varianza FAN	40
5.2.3.	Comparación de modelos fuera de muestra para Varianza FAN	40
5.2.4.	Entrenamiento modelos familia GARCH FAN (σ)	43
5.2.5.	Entrenamiento modelos híbridos Desviación Estándar FAN	43
5.2.6.	Comparación de modelos fuera de muestra para Desviación Estándar FAN	44
5.3.	ICLN	47
5.3.1.	Entrenamiento modelos familia GARCH ICLN (σ^2)	47
5.3.2.	Entrenamiento modelos híbridos Varianza ICLN	48
5.3.3.	Comparación de modelos fuera de muestra para Varianza ICLN	48
5.3.4.	Entrenamiento modelos familia GARCH ICLN (σ)	52
5.3.5.	Entrenamiento modelos híbridos Desviación Estándar ICLN	52
5.3.6.	Comparación de modelos fuera de muestra para Desviación Estándar ICLN	53
5.4.	IYE	56
5.4.1.	Entrenamiento modelos familia GARCH IYE (σ^2)	56
5.4.2.	Entrenamiento modelos híbridos Varianza IYE	57
5.4.3.	Comparación de modelos fuera de muestra para Varianza IYE	58
5.4.4.	Entrenamiento modelos familia GARCH IYE (σ)	61

5.4.5.	Entrenamiento modelos híbridos Desviación Estándar IYE	61
5.4.6.	Comparación de modelos fuera de muestra para Desviación Estándar IYE .	62
5.5.	PBW	65
5.5.1.	Entrenamiento modelos familia GARCH PBW (σ^2)	65
5.5.2.	Entrenamiento modelos híbridos Varianza PBW	66
5.5.3.	Comparación de modelos fuera de muestra para Varianza PBW	67
5.5.4.	Entrenamiento modelos familia GARCH PBW (σ)	70
5.5.5.	Entrenamiento modelos híbridos Desviación Estándar PBW	70
5.5.6.	Comparación de modelos fuera de muestra para Desviación Estándar PBW	71
5.6.	QCLN	75
5.6.1.	Entrenamiento modelos familia GARCH QCLN (σ^2)	75
5.6.2.	Entrenamiento modelos híbridos Varianza QCLN	76
5.6.3.	Comparación de modelos fuera de muestra para Varianza QCLN	77
5.6.4.	Entrenamiento modelos familia GARCH QCLN (σ)	80
5.6.5.	Entrenamiento modelos híbridos Desviación Estándar QCLN	80
5.6.6.	Comparación de modelos fuera de muestra para Desviación Estándar QCLN	81
5.7.	TAN	84
5.7.1.	Entrenamiento modelos familia GARCH TAN (σ^2)	84
5.7.2.	Entrenamiento modelos híbridos Varianza TAN	84
5.7.3.	Comparación de modelos fuera de muestra para Varianza TAN	86
5.7.4.	Entrenamiento modelos familia GARCH TAN (σ)	89
5.7.5.	Entrenamiento modelos híbridos Desviación estándar TAN	89
5.7.6.	Comparación de modelos fuera de muestra para Desviación Estándar TAN	90
5.8.	VDE	93
5.8.1.	Entrenamiento modelos familia GARCH VDE (σ^2)	93
5.8.2.	Entrenamiento modelos híbridos Varianza VDE	94
5.8.3.	Comparación de modelos fuera de muestra para Varianza VDE	95
5.8.4.	Entrenamiento modelos familia GARCH VDE (σ)	98
5.8.5.	Entrenamiento modelos Híbridos Desviación Estándar VDE	98

5.8.6.	Comparación de modelos fuera de muestra para Desviación Estándar VDE	99
5.9.	XLE	102
5.9.1.	Entrenamiento modelos familia GARCH XLE (σ^2)	102
5.9.2.	Entrenamiento modelos Híbridos Varianza XLE	102
5.9.3.	Comparación de modelos fuera de muestra para Varianza XLE	103
5.9.4.	Entrenamiento modelos familia GARCH XLE (σ)	107
5.9.5.	Entrenamiento modelos Híbridos Desviación Estándar XLE	107
5.9.6.	Comparación de modelos fuera de muestra para Desviación Estándar XLE	108
5.10.	XOP	111
5.10.1.	Entrenamiento modelos familia GARCH XOP (σ^2)	111
5.10.2.	Entrenamiento modelos Híbridos Varianza XOP	111
5.10.3.	Comparación de modelos fuera de muestra para Varianza XOP	112
5.10.4.	Entrenamiento modelos familia GARCH XOP (σ)	115
5.10.5.	Entrenamiento modelos Híbridos Desviación Estándar XOP	115
5.10.6.	Comparación de modelos fuera de muestra para Desviación Estándar XOP	115

6. Conclusiones	119
------------------------	------------

Índice de cuadros

1.	Estadísticas descriptivas ETFs energías limpias.	37
2.	Estadísticas descriptivas segunda mitad.	38
3.	Especificación de rezagos óptimos para modelos de familia GARCH FAN.	39
4.	Métricas de desempeño en entrenamiento familia GARCH FAN (σ^2).	39
5.	Hiperparámetros óptimos para varianza de FAN.	40
6.	Métricas de desempeño en entrenamiento FAN (σ^2).	40
7.	Resultados de métricas de desempeño FAN (σ^2).	41
8.	Conjunto superior de modelos para varianza de FAN.	41
9.	Métricas de desempeño en entrenamiento familia GARCH FAN (σ).	43
10.	Hiperparámetros óptimos para desviación estándar de FAN.	44
11.	Métricas de desempeño en entrenamiento modelos híbridos FAN (σ).	44
12.	Resultados de métricas de desempeño FAN (σ).	45
13.	Conjunto superior de modelos para desviación estándar de FAN.	45
14.	Especificación de rezagos óptimos para modelos de familia GARCH ICLN.	47
15.	Métricas de desempeño en entrenamiento familia GARCH ICLN (σ^2).	47
16.	Hiperparámetros óptimos para varianza de ICLN.	48
17.	Métricas de desempeño en entrenamiento modelos híbridos ICLN (σ^2).	49
18.	Resultados de métricas de desempeño ICLN (σ^2).	49
19.	Conjunto superior de modelos para varianza de ICLN.	50
20.	Métricas de desempeño en entrenamiento familia GARCH ICLN (σ).	52
21.	Hiperparámetros óptimos para desviación estándar de ICLN.	53
22.	Métricas de desempeño en entrenamiento modelos híbridos ICLN (σ).	53
23.	Resultados de métricas de desempeño ICLN (σ).	54
24.	Conjunto superior de modelos para desviación estándar de ICLN.	54
25.	Especificación de rezagos óptimos para modelos de familia GARCH IYE.	56
26.	Métricas de desempeño en entrenamiento familia GARCH IYE (σ^2).	57
27.	Hiperparámetros óptimos para varianza de IYE.	57

28.	Métricas de desempeño en entrenamiento modelos híbridos IYE (σ^2).	58
29.	Resultados de métricas de desempeño IYE (σ^2).	60
30.	Conjunto superior de modelos para varianza de IYE.	60
31.	Métricas de desempeño en entrenamiento familia GARCH IYE (σ).	61
32.	Hiperparámetros óptimos para desviación estándar de IYE.	62
33.	Métricas de desempeño en entrenamiento modelos híbridos IYE (σ).	62
34.	Resultados de métricas de desempeño IYE (σ).	63
35.	Conjunto superior de modelos para desviación estándar de IYE.	63
36.	Especificación de rezagos óptimos para modelos de familia GARCH PBW.	65
37.	Métricas de desempeño en entrenamiento familia GARCH PBW (σ^2).	66
38.	Hiperparámetros óptimos para varianza de PBW.	66
39.	Métricas de desempeño en entrenamiento modelos híbridos PBW (σ^2).	67
40.	Resultados de métricas de desempeño para la varianza de PBW (σ^2).	67
41.	Conjunto superior de modelos para varianza PBW.	68
42.	Métricas de desempeño en entrenamiento familia GARCH PBW (σ).	70
43.	Hiperparámetros óptimos para desviación estándar de PBW.	71
44.	Métricas de desempeño en entrenamiento modelos híbridos PBW (σ).	71
45.	Resultados de métricas de desempeño para el ETF PBW (σ).	72
46.	Conjunto superior de modelos para desviación estándar PBW.	74
47.	Especificación de rezagos óptimos para modelos de familia GARCH QCLN.	75
48.	Métricas de desempeño en entrenamiento familia GARCH QCLN (σ^2).	76
49.	Hiperparámetros óptimos para varianza de QCLN.	76
50.	Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ^2).	77
51.	Resultados de métricas de desempeño QCLN (σ^2).	79
52.	Conjunto superior de modelos para varianza de QCLN.	79
53.	Métricas de desempeño en entrenamiento familia GARCH QCLN (σ).	80
54.	Hiperparámetros óptimos para desviación estándar de QCLN.	81
55.	Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ).	81
56.	Resultados de métricas de desempeño QCLN (σ).	82

57.	Conjunto superior de modelos para desviación estándar de QCLN.	82
58.	Especificación de rezagos óptimos para modelos de familia GARCH TAN.	84
59.	Métricas de desempeño en entrenamiento familia GARCH TAN (σ^2).	84
60.	Hiperparámetros óptimos para varianza de TAN.	85
61.	Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ).	85
62.	Resultados de métricas de desempeño TAN (σ^2).	86
63.	Conjunto superior de modelos para varianza de TAN.	88
64.	Métricas de desempeño en entrenamiento familia GARCH TAN (σ).	89
65.	Hiperparámetros óptimos para desviación estándar de TAN.	90
66.	Métricas de desempeño en entrenamiento modelos híbridos TAN (σ).	90
67.	Resultados de métricas de desempeño TAN (σ).	91
68.	Conjunto superior de modelos para desviación estándar de TAN.	91
69.	Especificación de rezagos óptimos para modelos de familia GARCH VDE.	93
70.	Métricas de desempeño en entrenamiento familia GARCH VDE (σ^2).	94
71.	Hiperparámetros óptimos para varianza de VDE.	94
72.	Métricas de desempeño en entrenamiento modelos híbridos VDE (σ^2).	95
73.	Resultados de métricas de desempeño VDE (σ^2).	95
74.	Conjunto superior de modelos para varianza de VDE.	96
75.	Métricas de desempeño en entrenamiento familia GARCH VDE (σ).	98
76.	Hiperparámetros óptimos para desviación estándar de VDE.	99
77.	Métricas de desempeño en entrenamiento modelos híbridos VDE (σ).	99
78.	Resultados de métricas de desempeño VDE (σ).	99
79.	Conjunto superior de modelos para desviación estándar de VDE.	100
80.	Especificación de rezagos óptimos para modelos de familia GARCH XLE.	102
81.	Métricas de desempeño en entrenamiento familia GARCH XLE (σ^2).	102
82.	Hiperparámetros óptimos para varianza de XLE.	103
83.	Métricas de desempeño en entrenamiento modelos híbridos XLE (σ^2).	103
84.	Resultados de métricas de desempeño XLE (σ^2).	104
85.	Conjunto superior de modelos para varianza de XLE.	106

86.	Métricas de desempeño en entrenamiento familia GARCH XLE (σ).	107
87.	Hiperparámetros óptimos para desviación estándar de XLE.	107
88.	Métricas de desempeño en entrenamiento modelos híbridos XLE (σ).	108
89.	Resultados de métricas de desempeño XLE (σ).	108
90.	Conjunto superior de modelos para desviación estándar de XLE.	109
91.	Especificación de rezagos óptimos para modelos de familia GARCH XOP.	111
92.	Métricas de desempeño en entrenamiento familia GARCH XOP (σ^2).	111
93.	Hiperparámetros óptimos para varianza de XOP.	112
94.	Métricas de desempeño en entrenamiento modelos híbridos XOP (σ^2).	112
95.	Resultados de métricas de desempeño XOP (σ^2).	114
96.	Conjunto superior de modelos para varianza de XOP.	114
97.	Métricas de desempeño en entrenamiento familia GARCH XOP (σ).	115
98.	Hiperparámetros óptimos para desviación estándar de XOP.	116
99.	Métricas de desempeño en entrenamiento modelos híbridos XOP (σ).	116
100.	Resultados de métricas de desempeño XOP (σ).	116
101.	Conjunto superior de modelos para desviación estándar de XOP.	118

Índice de figuras

1.	Esquema de entrenamiento de un modelo híbrido LSTM–GARCH.	34
2.	Retornos Logarítmicos %	38
3.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF FAN (σ^2).	42
4.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF FAN (σ).	46
5.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF ICLN (σ^2).	51
6.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF ICLN (σ).	55
7.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF IYE (σ^2).	59
8.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF IYE (σ).	64
9.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF PBW (σ^2).	69
10.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF PBW (σ).	73

11.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF QCLN (σ^2).	78
12.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF QCLN (σ).	83
13.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF TAN (σ^2).	87
14.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF TAN (σ).	92
15.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF VDE (σ^2).	97
16.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF VDE (σ).	101
17.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XLE (σ^2).	105
18.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XLE (σ).	110
19.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XOP (σ^2).	113
20.	Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XOP (σ).	117

1. Problema de investigación

El presente estudio aborda la necesidad de modelar con precisión el fenómeno de la volatilidad en series de tiempo financieras correspondientes a fondos cotizados en bolsa (*Exchange Traded Funds*, ETFs) del mercado estadounidense, que replican índices de empresas vinculadas al sector energético, tanto de energías convencionales como de energías renovables.

La alta sensibilidad de estos instrumentos frente a factores macroeconómicos, crisis financieras y variaciones en los precios de los combustibles hace esencial el desarrollo de modelos que capturen adecuadamente su comportamiento dinámico y no lineal. En este contexto, surge la necesidad de comparar enfoques econométricos tradicionales, basados en modelos de la familia GARCH, con redes neuronales recurrentes de tipo LSTM, capaces de aprender dependencias complejas en el tiempo.

De este modo, la investigación busca aportar evidencia empírica sobre la efectividad de los modelos híbridos en la predicción diaria de la volatilidad, contribuyendo al entendimiento del riesgo financiero y ofreciendo información de valor para analistas e inversionistas del sector energético.

2. Objetivos de la investigación

2.1. Objetivo general

Evaluar y comparar el desempeño predictivo de modelos de volatilidad clásicos —pertenecientes a la familia GARCH— frente a modelos híbridos basados en redes neuronales LSTM, utilizando un conjunto de datos fuera de muestra que permita garantizar una comparación justa y metodológicamente robusta.

2.2. Objetivos específicos

1. Analizar las propiedades estadísticas y la estacionariedad de los retornos de los ETFs energéticos.
2. Optimizar y ajustar modelos de varianza condicional clásicos (GARCH, GJR-GARCH y EGARCH) mediante criterios de información.
3. Diseñar, entrenar y optimizar modelos híbridos LSTM–GARCH, integrando información econométrica y de aprendizaje profundo.
4. Incorporar la volatilidad implícita del índice VIX como variable exógena para evaluar su impacto en la predicción de la volatilidad.
5. Comparar el desempeño de los modelos en la predicción de la varianza y desviación estándar, tanto en ETFs de energías renovables como convencionales.
6. Determinar la robustez estadística de los resultados mediante el procedimiento de Model Confidence Set (MCS).

3. Marco Teórico

3.1. Fundamentos del Mercado Bursátil

Según la Comisión para el Mercado Financiero (**cmf**), el mercado bursátil se define como el “mercado secundario de compra y venta de valores de capitalización y de deuda”. Este mercado incluye diversos instrumentos financieros, como acciones, bonos y fondos cotizados en bolsa (ETFs), los cuales pueden representar tanto capital como deuda.

Estos instrumentos pueden tener un valor fiduciario —respaldado por la confianza en una entidad emisora— y/o un valor intrínseco, como ocurre en el caso de materias primas financieras como el oro o la plata.

El mercado bursátil cumple un rol fundamental en la economía, al permitir la canalización eficiente del ahorro hacia la inversión, facilitando la liquidez de los activos financieros y contribuyendo a la formación de precios mediante la interacción entre oferta y demanda (Grinblatt & Titman, 2011).

En este tipo de mercados, dos agentes económicos participan en la transacción de un activo financiero. Por ejemplo, en el caso de un bono emitido por la Tesorería de los Estados Unidos, la entidad emisora (la Tesorería) ofrece un compromiso de pago futuro más intereses anuales o semestrales, mientras que el inversionista entrega liquidez actual a cambio de dicho compromiso (Fabozzi, 2002).

3.2. Mercado Bursátil de Estados Unidos

El mercado bursátil estadounidense tiene sus orígenes a finales del siglo XVIII, cuando en 1792 veinticuatro corredores firmaron el Acuerdo de Buttonwood en Nueva York, sentando las bases de lo que más tarde se convertiría en la Bolsa de Valores de Nueva York (NYSE). Desde entonces, el mercado ha evolucionado hasta convertirse en uno de los sistemas financieros más desarrollados y líquidos del mundo.

Actualmente, los principales mercados bursátiles de Estados Unidos incluyen la NYSE y el

NASDAQ, donde se transan una amplia variedad de activos financieros. Entre ellos destacan las acciones de empresas públicas, bonos corporativos y gubernamentales, fondos cotizados en bolsa (ETFs), opciones, futuros y otros instrumentos derivados. Estos activos permiten a inversionistas e instituciones participar en la asignación de capital, la cobertura de riesgos y la especulación financiera (Mishkin, 2018).

3.3. Fondos Cotizables en Bolsa (ETFs) Energéticos

Los *Exchange Traded Funds* (ETFs) son vehículos de inversión colectiva que cotizan en bolsa y replican el desempeño de índices financieros específicos. A diferencia de los fondos mutuos tradicionales, los ETFs pueden comprarse y venderse durante toda la jornada bursátil, lo que los hace instrumentos líquidos, transparentes y de bajo costo para los inversionistas (Gastineau, 2010). En esencia, estos fondos permiten obtener exposición diversificada a un sector, región o clase de activo mediante una única transacción.

En el contexto del sector energético, los ETFs ofrecen una herramienta eficiente para comparar el desempeño de compañías asociadas tanto a las energías renovables como a las energías convencionales. Su estructura facilita la medición del riesgo y la rentabilidad relativa de cada segmento del mercado energético global, al reunir bajo un mismo instrumento empresas dedicadas a generación, exploración, distribución y tecnología energética.

A continuación, se describen los principales ETFs analizados en este estudio, de acuerdo con la información técnica y los prospectos de sus emisores oficiales:

- **First Trust Global Wind Energy ETF (FAN):** replica el *ISE Global Wind Energy Index*, compuesto por empresas internacionales dedicadas al desarrollo, fabricación e instalación de tecnologías eólicas, abarcando tanto generadores eléctricos como fabricantes de turbinas (First Trust Advisors, 2024a).
- **iShares Global Clean Energy ETF (ICLN):** invierte en una cartera diversificada de compañías del sector de energía limpia y tecnologías asociadas, incluyendo generación solar, eólica e hidrógeno. Su objetivo es seguir el desempeño del *S&P Global Clean Energy Index*

(BlackRock, Inc., 2024a).

- **iShares U.S. Energy ETF (IYE):** replica el *Russell 1000 Energy Index*, compuesto principalmente por grandes empresas estadounidenses de petróleo, gas y combustibles fósiles (BlackRock, Inc., 2024b).
- **Invesco WilderHill Clean Energy ETF (PBW):** enfocado en empresas estadounidenses vinculadas al desarrollo de tecnologías de energía limpia y reducción de emisiones de carbono. Su composición es altamente dinámica, reflejando la innovación del sector renovable (Invesco Ltd., 2024b).
- **First Trust NASDAQ Clean Edge Green Energy Index Fund (QCLN):** sigue el *NASDAQ Clean Edge Green Energy Index*, integrado por compañías de energías alternativas, almacenamiento y eficiencia energética, incluyendo fabricantes de vehículos eléctricos (First Trust Advisors, 2024b).
- **Invesco Solar ETF (TAN):** especializado en empresas dedicadas al sector solar fotovoltaico, abarcando productores de módulos, instaladores y proveedores de materiales clave. Representa uno de los fondos más sensibles a los ciclos de subsidios y políticas energéticas (Invesco Ltd., 2024a).
- **Vanguard Energy ETF (VDE):** busca replicar el *MSCI US Investable Market Energy 25/50 Index*, cubriendo una amplia gama de compañías estadounidenses de exploración, refinación y servicios petroleros. Constituye un referente del mercado energético convencional (Vanguard Group, Inc., 2024).
- **Energy Select Sector SPDR Fund (XLE):** agrupa a las principales empresas energéticas del índice *S&P 500*, incluyendo a ExxonMobil y Chevron. Es el ETF energético más líquido y representativo del mercado estadounidense (State Street Global Advisors, 2024a).
- **SPDR S&P Oil & Gas Exploration & Production ETF (XOP):** enfocado en empresas dedicadas a la exploración y producción de petróleo y gas natural en Estados Unidos. Su cartera pondera equitativamente las compañías, reflejando de mejor forma la evolución del

sector *upstream* (State Street Global Advisors, 2024b).

La selección de estos nueve ETFs se justifica por los criterios de alta liquidez y disponibilidad de datos históricos consistentes desde 2008, y su representatividad de los sectores de energía limpia y fósil dentro del mercado estadounidense y global.

3.4. Índice de Volatilidad (VIX)

El índice VIX, desarrollado por el Chicago Board Options Exchange (CBOE), mide la expectativa implícita de volatilidad a 30 días del índice S&P 500 a partir de precios de opciones. Es conocido comúnmente como el “índice del miedo”, dado que tiende a incrementarse en períodos de incertidumbre financiera, reflejando la aversión al riesgo de los inversionistas (Whaley, 2000).

3.5. Modelos de Varianza Condicional

Los modelos de varianza condicional, iniciados con la propuesta ARCH de Engle (1982), constituyen una de las herramientas más utilizadas en econometría financiera para modelar la volatilidad. Su principal aporte consiste en capturar la heterocedasticidad condicional, es decir, la dependencia de la varianza de los retornos respecto a choques pasados. Posteriormente, Bollerslev (1986) extendió este marco al proponer el modelo GARCH, lo que permitió describir de forma más parsimoniosa la persistencia de la volatilidad. Otras extensiones relevantes son el EGARCH (Nelson, 1991) y el GJR-GARCH (Glosten et al., 1993), que permiten modelar las asimetrías observadas en los mercados financieros.

En todos estos casos, la estimación de parámetros se lleva a cabo mediante el método de máxima verosimilitud (MLE). Bajo el supuesto de errores normalmente distribuidos, la log-verosimilitud general para n observaciones se expresa como:

$$\ell(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \left[\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2} \right]$$

donde σ_t^2 es la varianza condicional y θ el vector de parámetros del modelo. Este procedi-

miento garantiza estimadores consistentes y eficientes cuando se cumplen los supuestos básicos.

3.5.1. Modelo ARCH

El modelo ARCH (*Autoregressive Conditional Heteroskedasticity*) fue introducido por Engle (1982), y constituye el punto de partida de la modelación moderna de la volatilidad financiera. Este modelo asume que la varianza condicional de los retornos depende de los errores pasados, capturando así el fenómeno de *volatility clustering*, donde periodos de alta y baja volatilidad tienden a agruparse temporalmente.

La especificación básica se compone de una ecuación de media y una de varianza condicional. La media puede ser constante o modelarse mediante un proceso ARMA, mientras que la varianza depende de los choques pasados:

$$r_t = \bar{r} + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

donde r_t corresponde al retorno en el tiempo t , $\bar{r}$ es la media del proceso, ε_t representa el error y e_t es un término de ruido blanco. Aunque tradicionalmente se asume que $e_t \sim N(0, 1)$, esta hipótesis no es estrictamente necesaria: las distribuciones no normales (por ejemplo, t de Student o GED) suelen capturar mejor las colas pesadas observadas empíricamente en los retornos financieros.

3.5.2. Modelo GARCH

El modelo GARCH (*Generalized ARCH*) propuesto por Bollerslev (1986) amplía el modelo de Engle al incluir rezagos autorregresivos de la varianza condicional. Esto permite una representación más parsimoniosa y eficiente de la persistencia en la volatilidad:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

El caso más utilizado, GARCH(1,1), suele capturar adecuadamente la dinámica de la volatilidad de activos financieros, al reflejar el alto grado de memoria de los choques en la varianza. Nuevamente, la normalidad condicional no es un supuesto indispensable; los modelos pueden especificarse bajo distribuciones más flexibles para los errores, mejorando la estimación de riesgos extremos.

3.5.3. Modelo EGARCH

El modelo EGARCH (*Exponential GARCH*), desarrollado por Nelson (1991), introduce una formulación logarítmica para garantizar la positividad de la varianza condicional sin imponer restricciones sobre los parámetros. Además, permite modelar la asimetría en la respuesta de la volatilidad ante choques positivos o negativos, fenómeno conocido como *efecto apalancamiento*:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^p \alpha_i \left(\frac{|\varepsilon_{t-i}|}{\sigma_{t-i}} - \sqrt{\frac{2}{\pi}} \right) + \sum_{k=1}^o \gamma_i \frac{\varepsilon_{t-k}}{\sigma_{t-k}} + \sum_{j=1}^q \beta_j \ln(\sigma_{t-j}^2)$$

El parámetro γ_i mide el grado de asimetría: si $\gamma_i < 0$, los choques negativos incrementan en mayor medida la volatilidad que los positivos de igual magnitud, reflejando la respuesta típica de los mercados financieros ante noticias adversas.

3.5.4. Modelo GJR-GARCH

El modelo GJR-GARCH, propuesto por Glosten et al. (1993), constituye una alternativa al EGARCH para capturar la asimetría de los retornos mediante una variable indicadora:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{k=1}^o \gamma_i I_{\{\varepsilon_{t-k} < 0\}} \varepsilon_{t-k}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

donde $I_{\{\varepsilon_{t-i} < 0\}}$ toma el valor 1 si el choque es negativo y 0 en caso contrario. Si $\gamma_i > 0$, se confirma la presencia del efecto apalancamiento. Este modelo es ampliamente usado en la práctica por su simplicidad y capacidad para capturar asimetrías sin requerir transformaciones logarítmicas.

3.6. Modelos GARCH–VIX

Los modelos GARCH–VIX surgen como una extensión natural de los modelos GARCH tradicionales al incorporar información exógena proveniente del índice de volatilidad implícita del mercado (VIX).

De acuerdo con Gutiérrez (2023), la inclusión del VIX dentro de la ecuación de varianza condicional permite capturar una componente de volatilidad anticipada que no está contenida en los choques históricos de los precios. Esta aproximación mejora la capacidad predictiva de los modelos GARCH, al integrar en la estructura de la varianza información proveniente del mercado de opciones.

Formalmente, la especificación del modelo GARCH–VIX puede expresarse como:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{k=1}^z \theta_k VIX_{t-k}$$

y su versión asimétrica (GJR–GARCH–VIX) incorpora el efecto apalancamiento mediante un término adicional que se activa ante choques negativos:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^o \gamma_i I_{\{\varepsilon_{t-i} < 0\}} \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{k=1}^z \theta_k VIX_{t-k}$$

En este contexto, el parámetro θ_k mide el impacto de la volatilidad implícita sobre la volatilidad condicional. Un valor positivo indica que un aumento del VIX se traduce en una mayor varianza esperada, coherente con el rol del índice como indicador de riesgo sistémico.

Aunque se evaluó también el modelo EGARCH–VIX, sus resultados fueron inestables: a pesar de alcanzar convergencia numérica, las estimaciones arrojaron errores del orden de 10^{30} , motivo por el cual se descartó su uso en este estudio.

3.7. Machine Learning

El campo de *Machine Learning* (ML) constituye una rama fundamental de la inteligencia artificial, cuyo propósito es desarrollar algoritmos y modelos capaces de aprender patrones a partir de datos y mejorar su desempeño de manera autónoma. Una definición ampliamente citada es la propuesta por Mitchell (1997), quien establece que: “Se dice que un programa computacional aprende de la experiencia E , respecto de una tarea T y una medida de desempeño P , si el desempeño en T , medido mediante P , mejora con la experiencia E ”.

El aprendizaje automático ha demostrado ser una herramienta esencial en múltiples disciplinas, tales como las finanzas, la ingeniería, la medicina y las ciencias sociales, gracias a su capacidad para procesar grandes volúmenes de información y capturar relaciones complejas que los métodos estadísticos tradicionales no siempre pueden modelar (Goodfellow et al., 2016; LeCun et al., 2015).

Existen diferentes tipos de algoritmos y enfoques en machine learning, los cuales suelen clasificarse de la siguiente manera:

- **Aprendizaje supervisado:** el modelo se entrena con datos etiquetados, aprendiendo a predecir una salida a partir de una entrada. Ejemplos comunes incluyen la regresión lineal, máquinas de soporte vectorial (*Support Vector Machines*) y redes neuronales profundas.
- **Aprendizaje no supervisado:** los datos no poseen etiquetas, y el objetivo es descubrir estructuras ocultas, como agrupamientos o reducciones de dimensionalidad. Ejemplos típicos son el algoritmo k -means o el análisis de componentes principales (PCA).
- **Aprendizaje por refuerzo:** el agente aprende a partir de la interacción con un entorno, recibiendo recompensas o penalizaciones en función de sus acciones. Este enfoque ha sido clave en aplicaciones de robótica y en sistemas de decisión secuenciales.
- **Aprendizaje en línea (*online learning*):** el modelo se actualiza de manera continua conforme recibe nuevas observaciones, lo que lo hace especialmente adecuado para entornos dinámicos como los mercados financieros.

El presente estudio pone especial énfasis en los modelos basados en redes neuronales re-

corrientes (RNN) y sus variantes más avanzadas, particularmente las *Long Short-Term Memory* (LSTM), dada su capacidad para modelar series de tiempo no lineales y dependencias de largo plazo, características propias de los datos financieros.

3.7.1. Redes Neuronales Recurrentes (RNN)

Las redes neuronales recurrentes introducen conexiones de retroalimentación que permiten mantener un estado interno dependiente de las entradas previas. Esta propiedad las convierte en modelos idóneos para secuencias temporales, ya que capturan dependencias dinámicas entre observaciones (Rumelhart et al., 1986). Sin embargo, presentan limitaciones relacionadas con el *desvanecimiento* y la *explosión* del gradiente, lo cual dificulta el aprendizaje de relaciones a largo plazo.

3.7.2. Redes LSTM

Las redes LSTM (*Long Short-Term Memory*), propuestas por Hochreiter y Schmidhuber (1997), surgen como una extensión de las RNN convencionales. Su principal innovación radica en la introducción de una célula de memoria regulada por compuertas (*gates*) que controlan el flujo de información. Gracias a este diseño, las LSTM superan el problema del desvanecimiento del gradiente y pueden aprender dependencias de largo alcance en secuencias temporales. Esto las ha posicionado como un modelo de referencia en tareas de predicción de series de tiempo, reconocimiento de voz y procesamiento de lenguaje natural (Graves, 2013). Una LSTM introduce una estructura denominada “célula de memoria”, que actúa como un “contenedor” capaz de mantener información a lo largo del tiempo. Esta célula está controlada por tres compuertas (*gates*), cada una de las cuales regula el flujo de información:

- Input gate (i_t): decide qué parte de la nueva información de la entrada x_t se incorporará al estado de la célula.
- Forget gate (f_t): determina qué parte de la información previa almacenada en la memoria debe olvidarse.
- Output gate (o_t): regula qué parte de la información contenida en la célula se transmite como

salida h_t .

Adicionalmente, se introduce una candidata de estado de célula (g_t), que representa nueva información que podría incorporarse a la memoria.

La dinámica de una célula LSTM se describe mediante las siguientes ecuaciones:

$$i_t = \sigma(W_{xi}^T x_t + W_{hi}^T h_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}^T x_t + W_{hf}^T h_{t-1} + b_f)$$

$$o_t = \sigma(W_{xo}^T x_t + W_{ho}^T h_{t-1} + b_o)$$

$$g_t = \tanh(W_{xg}^T x_t + W_{hg}^T h_{t-1} + b_g)$$

El estado de la célula (c_t) se actualiza combinando el estado previo y la información nueva:

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t$$

Finalmente, la salida de la célula (h_t) se obtiene aplicando la compuerta de salida al estado actualizado:

$$h_t = o_t \otimes \tanh(c_t)$$

En ese esquema, $\sigma(\cdot)$ representa la función sigmoide, que restringe los valores entre 0 y 1 (control de flujo), $\tanh(\cdot)$ es la función tangente hiperbólica, y $\otimes$ denota el producto elemento a elemento.

3.7.3. Modelos Híbridos LSTM-GARCH

Los modelos híbridos que combinan técnicas econométricas con redes neuronales artificiales tienen su origen en el trabajo pionero de Donaldson y Kamstra (1997), quienes integraron un modelo ARCH con una red neuronal de tipo perceptrón multicapa. Su estudio demostró mejoras

sustanciales en la predicción de la volatilidad de índices bursátiles de Estados Unidos, Japón y Europa, en comparación con los modelos GARCH tradicionales.

En las últimas décadas, el desarrollo de capacidades computacionales y el avance del aprendizaje profundo han impulsado una nueva generación de modelos híbridos que combinan estructuras GARCH con redes neuronales recurrentes, particularmente del tipo LSTM (*Long Short-Term Memory*). Estos modelos, introducidos inicialmente en el ámbito financiero por Bao et al. (2017), han mostrado un desempeño superior al de los enfoques puramente econométricos o puramente neuronales.

La motivación detrás de esta arquitectura radica en aprovechar las fortalezas de ambos paradigmas: los modelos GARCH capturan la dinámica de la varianza condicional y la persistencia temporal de la volatilidad, mientras que las redes LSTM son capaces de modelar relaciones no lineales y dependencias de largo plazo en las series financieras. En conjunto, los modelos híbridos LSTM–GARCH han demostrado mejoras significativas en la predicción de volatilidad, reducción del error cuadrático medio y una mayor estabilidad en contextos de alta incertidumbre del mercado.

3.7.4. Modelos Híbridos LSTM–GARCH–VIX

Una extensión reciente consiste en la incorporación del índice VIX como variable exógena dentro de los modelos híbridos LSTM–GARCH. Este enfoque, conocido como LSTM–GARCH–VIX, busca integrar no solo la estructura de heterocedasticidad condicional de los retornos y la capacidad de las LSTM para aprender patrones secuenciales, sino también la información de expectativas de volatilidad del mercado (Mutinda & Langat, 2024). En consecuencia, estos modelos permiten capturar de manera más realista la interacción entre factores internos y externos en la dinámica de los precios financieros.

3.8. Optimizador Bayesiano

La selección de hiperparámetros es un aspecto crítico en el entrenamiento de modelos de *machine learning*. Los métodos tradicionales, como la búsqueda en malla (*grid search*) o la búsqueda aleatoria (*random search*), suelen ser ineficientes cuando el espacio de hiperparámetros es amplio

y costoso de evaluar. Frente a estas limitaciones, la optimización bayesiana surge como una estrategia más eficiente, especialmente en contextos donde cada evaluación del modelo implica un costo computacional significativo (Shahriari et al., 2016).

La optimización bayesiana construye un modelo probabilístico del desempeño de la función objetivo $f(\mathbf{x})$ sobre el espacio de hiperparámetros $\mathbf{x} \in \mathcal{X}$. Generalmente, este modelo se implementa como un proceso gaussiano (*Gaussian Process*) o como un modelo basado en árboles (*Tree-structured Parzen Estimator*, TPE). En lugar de evaluar directamente todos los puntos, el optimizador utiliza una función de adquisición que equilibra exploración y explotación para decidir qué conjunto de hiperparámetros probar en la siguiente iteración. Formalmente, se busca:

$$\mathbf{x}_{t+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} a(\mathbf{x} | D_t)$$

donde $a(\mathbf{x} | D_t)$ es la función de adquisición condicionada en los datos observados $D_t = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^t$.

En particular, la librería *Optuna* implementa el método TPE, el cual modela la distribución condicional de los hiperparámetros dado el valor de la función objetivo. En este enfoque, se construyen dos densidades: una para los conjuntos de hiperparámetros que conducen a buenos resultados y otra para los que conducen a peores resultados. El optimizador selecciona nuevas configuraciones $\mathbf{x}$ maximizando la relación de verosimilitud entre ambas distribuciones, lo que permite explorar de manera eficiente regiones prometedoras del espacio de búsqueda (Akiba et al., 2019).

3.9. Métricas de Evaluación

La comparación y validación de modelos econométricos y de machine learning requiere el uso de métricas de evaluación que permitan cuantificar su desempeño. En este estudio se emplean criterios de información, medidas de error y tests de comparación estadística, los cuales se describen a continuación:

- **Akaike Information Criterion (AIC):** propuesto por Akaike (1974), el AIC permite selec-

cionar entre modelos rivales considerando tanto la bondad de ajuste como la complejidad del modelo. Se define como:

$$AIC = -2\ell(\hat{\theta}) + 2k$$

donde $\ell(\hat{\theta})$ corresponde al valor de la log-verosimilitud en el óptimo y k al número de parámetros estimados. Modelos con menor AIC son preferidos, ya que logran un equilibrio entre ajuste y parsimonia.

- **Mean Squared Error (MSE):** es una de las métricas más utilizadas para cuantificar errores de predicción. Se define como el promedio de los errores al cuadrado:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

donde y_t es el valor real y $\hat{y}_t$ la predicción. Penaliza más fuertemente los errores grandes, lo que lo hace adecuado en contextos donde estos deben evitarse.

- **Coefficiente de determinación (R^2):** mide la proporción de la variabilidad observada en la variable dependiente que es explicada por el modelo. Se interpreta como una medida de bondad de ajuste, donde valores cercanos a 1 indican un modelo con alta capacidad explicativa. Matemáticamente, se define como:

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2},$$

donde y_t representa los valores observados, $\hat{y}_t$ los valores estimados por el modelo y $\bar{y}$ la media muestral de y_t .

- **Mean Absolute Error (MAE):** mide el error promedio absoluto entre valores observados y predichos:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

A diferencia del MSE, todos los errores tienen el mismo peso, lo que lo hace más robusto frente a valores atípicos.

- **Heteroskedasticity-adjusted Mean Squared Error (HMSE):** introducido por Fuertes et al. (2009), ajusta el MSE para tener en cuenta la heterocedasticidad presente en series financieras. Al ponderar los errores por la varianza condicional estimada, permite una evaluación más realista del desempeño de modelos de volatilidad.

$$HMSE = \frac{1}{n} \sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{\hat{\sigma}_t^2}$$

donde $\hat{\sigma}_t^2$ representa la varianza condicional estimada.

- **Heteroskedasticity-adjusted Mean Absolute Error (HMAE):** de forma análoga, el HMAE extiende el MAE ajustando los errores absolutos por la varianza condicional. Es particularmente útil en entornos financieros donde la volatilidad varía a lo largo del tiempo.

$$HMAE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\hat{\sigma}_t}$$

- **Mean Directional Accuracy (MDA):** mide la proporción de ocasiones en que el modelo predice correctamente la dirección del cambio de una variable financiera, más allá de su magnitud. Esta métrica es especialmente relevante en el contexto de predicción de retornos o volatilidad, donde acertar el sentido del movimiento resulta más importante que estimar su valor exacto.

$$MDA = \frac{1}{T} \sum_{t=1}^T I[(r_t - r_{t-1})(\hat{r}_t - r_{t-1}) > 0]$$

donde:

- r_t representa el valor observado en el tiempo t ,
- $\hat{r}_t$ es el valor predicho,
- $I(\cdot)$ es la función indicadora que toma valor 1 si ambos términos tienen el mismo signo y 0 en caso contrario,

- T es el número total de observaciones.

El valor de la MDA se encuentra en el rango $[0, 1]$. Un valor cercano a 1 indica una alta precisión direccional, mientras que valores alrededor de 0,5 sugieren predicciones aleatorias, y valores menores a 0,5 implican una direccionalidad inversa. Esta métrica fue utilizada por Pesaran y Timmermann (1992) como una herramienta robusta para evaluar la capacidad predictiva no paramétrica de los modelos financieros.

- **Quasi-Likelihood Loss (QLIKE)** es una métrica utilizada por Patton (2011) para evaluar la precisión de modelos de predicción de volatilidad, con proxis de la misma. Su principal ventaja es que penaliza de manera asimétrica los errores de sobreestimación y subestimación, otorgando mayor peso a las observaciones donde la volatilidad real es alta, lo que la hace especialmente adecuada para modelos GARCH y sus extensiones.

$$QLIKE = \frac{1}{T} \sum_{t=1}^T \left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} - \ln \frac{\sigma_t^2}{\hat{\sigma}_t^2} - 1 \right)$$

donde:

- σ_t^2 representa la varianza condicional observada (realizada),
- $\hat{\sigma}_t^2$ corresponde a la varianza condicional estimada por el modelo.

Un valor menor de QLIKE indica un mejor ajuste del modelo. Dado que la función de pérdida es insensible a transformaciones monotónicas, su uso es estándar en la literatura de comparación de pronósticos de volatilidad (Hansen & Lunde, 2005).

- **Model Confidence Set (MCS):** propuesto por Hansen et al. (2011), es un procedimiento estadístico diseñado para comparar el desempeño predictivo de un conjunto de modelos alternativos bajo un marco de inferencia múltiple. A diferencia de los enfoques tradicionales que seleccionan un único modelo en función de un criterio de información o de error, el MCS identifica un *conjunto de modelos estadísticamente equivalentes* en términos de desempeño, dentro de un nivel de confianza predefinido $(1 - \alpha)$.

El procedimiento MCS se basa en pruebas de hipótesis secuenciales que contrastan la igualdad de desempeño entre modelos a partir de una función de pérdida $L_{i,t}$ (por ejemplo, el error cuadrático medio, MSE , o el $QLIKE$). La diferencia de desempeño entre dos modelos i y j se define como:

$$d_{ij,t} = L_{i,t} - L_{j,t}$$

y su media muestral como:

$$\bar{d}_{ij} = \frac{1}{T} \sum_{t=1}^T d_{ij,t}$$

Las hipótesis nula y alternativa se plantean de la siguiente manera:

$$H_0 : E(d_{ij}) = 0 \quad \forall i, j \in M \quad H_1 : \exists (i, j) \text{ tal que } E(d_{ij}) \neq 0$$

donde M representa el conjunto de modelos bajo evaluación.

Para contrastar estas hipótesis, se construyen estadísticos de prueba basados en la varianza de $\bar{d}_{ij}$, la cual se estima mediante un procedimiento de *bootstrap bloqueado* que preserva la dependencia temporal de los datos. Este enfoque permite obtener distribuciones empíricas robustas frente a autocorrelación o heterocedasticidad en los errores de predicción.

Los dos estadísticos más comunes empleados son:

$$T_{\max} = \max_{i,j \in M} \frac{|\bar{d}_{ij}|}{\hat{\sigma}(\bar{d}_{ij})} \quad \text{y} \quad T_R = \max_{i \in M} t_i,$$

donde t_i corresponde al promedio estandarizado de las pérdidas relativas del modelo i respecto al resto.

El procedimiento MCS elimina iterativamente el modelo con peor desempeño promedio hasta que no se pueda rechazar la hipótesis nula de igualdad de desempeño entre los modelos restantes. El conjunto resultante se denomina *Model Confidence Set* con nivel de confianza $(1 - \alpha)$.

En este estudio, se utiliza un nivel de significancia $\alpha = 0,05$ y 10.000 replicaciones de boots-

trap para estimar la distribución del estadístico $T_{\text{máx}}$. De esta forma, el MCS permite identificar no solo el mejor modelo individual, sino un subconjunto de modelos con desempeño estadísticamente indistinguible, proporcionando un marco de comparación más robusto ante la incertidumbre muestral y el sobreajuste.

4. Metodología

El presente estudio se desarrolla a partir de series de tiempo financieras correspondientes a los precios de cierre diarios de distintos ETFs del sector energético, obtenidos desde fuentes abiertas como *Yahoo Finance*. Dado que se trabaja con datos de baja frecuencia, el precio de cierre en el día t se denota como C_t , y su retorno logarítmico porcentual se define como:

$$r_t \% = \ln \left(\frac{C_t}{C_{t-1}} \right) \cdot 100$$

De esta manera, la aproximación más directa a la varianza y desviación estándar diaria de los retornos se expresa como:

$$\sigma_t^2 = r_t^2 \quad \text{y} \quad \sigma_t = |r_t|$$

Estas magnitudes permiten cuantificar la volatilidad de corto plazo en las series bajo estudio y constituyen las variables dependientes en los modelos de predicción.

4.1. Análisis Exploratorio

Se efectúa un análisis exploratorio de los retornos porcentuales con el objetivo de identificar patrones, propiedades estadísticas y posibles valores atípicos. Este proceso incluye el cálculo de métricas de primer y segundo orden, así como la aplicación de pruebas de hipótesis orientadas a verificar los supuestos fundamentales del modelado de series financieras.

Entre las medidas descriptivas consideradas destacan:

- **Media:** representa el rendimiento promedio esperado.

$$\bar{r} = \frac{1}{n} \sum_{t=1}^n r_t$$

- **Desviación estándar:** cuantifica la dispersión de los retornos en torno a la media.

$$s = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (r_t - \bar{r})^2}$$

- **Asimetría (Skewness):** mide la simetría de la distribución de retornos. Valores negativos indican una mayor probabilidad de pérdidas extremas.

$$S = \frac{1}{n} \sum_{t=1}^n \left(\frac{r_t - \bar{r}}{s} \right)^3$$

- **Curtosis:** evalúa el grado de concentración de la distribución en torno a la media. Valores superiores a 3 implican colas pesadas (*leptocúrticas*), típicas en activos financieros.

$$K = \frac{1}{n} \sum_{t=1}^n \left(\frac{r_t - \bar{r}}{s} \right)^4$$

Además, se aplican diversas pruebas estadísticas para contrastar los supuestos de normalidad, estacionariedad, independencia y presencia de heterocedasticidad condicional en los retornos:

- **Prueba de Jarque–Bera (JB):** evalúa la hipótesis de normalidad de la distribución de los retornos, combinando los momentos de asimetría y curtosis (Jarque y Bera (1980)).

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right)$$

Hipótesis:

H_0 : Los retornos siguen una distribución normal. H_1 : Los retornos no siguen una distribución normal.

Si $JB > \chi_{2,0,05}^2 = 5,99$, se rechaza H_0 al 5 %.

- **Prueba Dickey–Fuller Aumentada (ADF):** contrasta la hipótesis nula de existencia de raíz unitaria (Dickey y Fuller (1979)). La regresión considerada es:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^k \delta_i \Delta y_{t-i} + \varepsilon_t$$

Hipótesis:

$$H_0 : \gamma = 0 \quad (\text{la serie no es estacionaria}) \quad H_1 : \gamma < 0 \quad (\text{la serie es estacionaria})$$

Si el estadístico ADF es menor que el valor crítico (−3,43 al 1 %, −2,86 al 5 %, −2,57 al 10 %).

- **Prueba de Phillips–Perron (PP):** alternativa robusta a la ADF, corrige por autocorrelación y heterocedasticidad de los errores (Phillips y Perron (1988)). Utiliza un estadístico similar al ADF:

$$Z_{t_\alpha} = t_\alpha - \frac{\hat{s}^2 - \hat{\sigma}^2}{2\hat{\sigma}^2} \left[\frac{\sum_{t=1}^T y_{t-1}^2}{T} \right]$$

Hipótesis:

$$H_0 : \text{la serie presenta raíz unitaria} \quad H_1 : \text{la serie es estacionaria}$$

Al igual que en la ADF, valores del estadístico menores a los críticos indican estacionariedad.

- **Prueba KPSS (Kwiatkowski–Phillips–Schmidt–Shin):** verifica la estacionariedad bajo una hipótesis nula opuesta a la ADF (Kwiatkowski et al. (1992)). El estadístico se define como:

$$\eta = \frac{1}{T^2 \hat{\sigma}^2} \sum_{t=1}^T S_t^2, \quad S_t = \sum_{i=1}^t \hat{u}_i$$

donde $\hat{u}_i$ son los residuos de una regresión lineal de y_t sobre una constante o una tendencia.

Hipótesis:

$$H_0 : \text{la serie es estacionaria} \quad H_1 : \text{la serie es no estacionaria}$$

Si $\eta > 0,739$ (1 %), 0,463 (5 %) o 0,347 (10 %), se rechaza H_0 . En este estudio, los valores de η oscilan entre 0.09 y 0.40, dentro de los límites de aceptación, por lo que no se rechaza la estacionariedad.

- **Prueba de Ljung–Box Q:** detecta autocorrelación en los residuos hasta un número m de rezagos (Ljung y Box (1978)). El estadístico se define como:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

donde $\hat{\rho}_k$ es la autocorrelación muestral en el rezago k .

Hipótesis:

$$H_0 : \text{no existe autocorrelación en los retornos.} \quad H_1 : \text{existe autocorrelación en los retornos.}$$

Si $Q > \chi_{m,0,05}^2$ o el p-valor es menor a 0.05, se rechaza H_0 .

En conjunto, las pruebas demostrarán la idoneidad de aplicar los modelos propuestos.

4.2. División de Datos

El estado del arte en el modelado de volatilidad en mercados energéticos recomienda particionar los datos en proporciones cercanas al 70–80 % para entrenamiento y 20–30 % para prueba (Zhang et al., 2019). En este estudio, el período analizado abarca desde el 01-07-2008 hasta el 30-05-2025. Se utiliza un 75 % de observaciones para el entrenamiento y un 25 % para la evaluación fuera de muestra (*out-of-sample*).

4.3. Entrenamiento de Modelos Clásicos

Se consideran tres especificaciones de la familia ARCH: GARCH, GJR-GARCH y EGARCH, ampliamente utilizadas para capturar persistencia y asimetría en la volatilidad de los retornos financieros. El ajuste se realiza sobre la muestra de entrenamiento, utilizando los retornos logarítmicos porcentuales. Esta transformación estabiliza la varianza y mejora la convergencia de los algoritmos de optimización.

Los parámetros se estiman mediante el método de máxima verosimilitud (MLE) con el algoritmo BFGS (*Broyden–Fletcher–Goldfarb–Shanno*), garantizando estimadores consistentes y eficientes bajo normalidad condicional. Se verifican las condiciones de positividad y estacionariedad de la varianza condicional en todos los modelos.

4.4. Entrenamiento de Modelos Híbridos

Los modelos híbridos combinan la estructura econométrica de los modelos GARCH con la capacidad de aprendizaje no lineal y secuencial de las redes neuronales LSTM. Este enfoque busca integrar información econométrica con aprendizaje profundo, mejorando la capacidad predictiva de la volatilidad bajo escenarios de alta complejidad.

4.4.1. Estandarización de Variables

Cada variable de entrada a la red ($\hat{\sigma}^2$, σ^2 , VIX) se somete a un proceso de normalización *min-max*, ajustado al rango $[-1, 1]$, calculado de forma independiente para cada variable. Este procedimiento garantiza escalas comparables, evita dominancia numérica y mejora la estabilidad del entrenamiento.

4.4.2. Entrenamiento Supervisado

La estructura de entrada preserva la causalidad temporal. Las variables σ^2 y VIX se incluyen con un rezago de un día para respetar la causalidad de Granger, mientras que la varianza condicional estimada ($\hat{\sigma}^2$) se incorpora de forma contemporánea con la variable objetivo, permitiendo que la

red aprenda a corregir errores sistemáticos de los modelos clásicos.

La Figura 1 ilustra el flujo general del proceso de entrenamiento de los modelos híbridos.



Figura 1: Esquema de entrenamiento de un modelo híbrido LSTM–GARCH.

El entrenamiento se realiza bajo validación cruzada con $k = 8$ pliegues, reduciendo el riesgo de sobreajuste y mejorando la robustez de las estimaciones. El criterio de optimización corresponde al error cuadrático medio (MSE) promedio entre pliegues. El optimizador empleado es Adam, debido a su eficiencia en el ajuste dinámico de la tasa de aprendizaje.

Se implementa un balanceo de cuartiles en la generación de *batches*, garantizando que cada lote contenga observaciones de diferentes niveles de volatilidad (baja, media y alta), lo que mejora la capacidad de generalización del modelo.

La arquitectura considera tres capas LSTM totalmente conectadas seguidas de una capa densa de salida, configuradas para capturar dependencias temporales complejas y generar estimaciones puntuales de volatilidad.

4.5. Optimización de Modelos Clásicos

La selección del modelo óptimo dentro de cada familia ARCH se realiza mediante un análisis sistemático de rezagos. Se construye una malla con combinaciones de p y q hasta cinco rezagos,

evaluando cada especificación mediante el criterio de información de Akaike (AIC). El modelo que minimiza el AIC se selecciona por ofrecer el mejor equilibrio entre ajuste y parsimonia.

4.6. Optimización de Modelos Híbridos

La búsqueda de hiperparámetros se lleva a cabo con *Optuna*, un marco de optimización bayesiana basado en el estimador *Tree-structured Parzen Estimator (TPE)*. Se definen espacios de búsqueda para: número de rezagos (1–15), neuronas por capa (2–40 o hasta 60 si se incluye VIX), tamaño de lote (30–70), descarte aleatorio (0–0,4) y tasa de aprendizaje (10^{-5} – 10^{-2}). Se ejecutan 50 iteraciones por modelo, generando un total de 250 configuraciones por serie y variable (σ o σ^2), priorizando configuraciones con menor MSE promedio.

4.7. Evaluación y Comparación

La evaluación final se realiza sobre el conjunto de prueba (*out-of-sample*). Los modelos se entrenan con datos hasta el 05-03-2021 y generan predicciones recursivas día a día sin reentrenamiento, simulando un entorno de predicción real.

Las métricas de desempeño empleadas son R^2 %, MSE, MAE, HMSE, HMAE, QLIKE, MDA y MCS, permitiendo evaluar precisión, consistencia y robustez estadística. Estos indicadores facilitan una comparación integral entre las familias de modelos, identificando aquellos con mejor capacidad predictiva de la volatilidad en los mercados energéticos.

5. Resultados

5.1. Análisis Exploratorio

El análisis descriptivo de las Tablas 1 y 2 de los ETFs (de aquí en adelante, se hará referencia a ellos por su nombre clave de transacción) revela patrones diferenciados en el comportamiento de los instrumentos financieros del sector energético. La primera agrupación, compuesta por FAN, TAN, PBW, ICLN y QCLN —índices asociados a empresas del rubro de las energías limpias—,

exhibe un mayor riesgo promedio como bloque en comparación con la segunda cohorte, integrada por IYE, VDE, XLE y XOP, pertenecientes a energías convencionales.

En términos de rentabilidad promedio, los ETFs del primer grupo presentan medias cercanas a la neutralidad, con valores que oscilan entre $-0,04176\%$ para TAN y $0,0069\%$ para QCLN. Destaca que, a pesar de estas medias modestas, todas las medianas son neutras o positivas y, en la mayoría de los casos, superiores a las medias, especialmente en QCLN ($0,1063\%$) y FAN ($0,0468\%$). Este patrón sugiere la presencia de distribuciones con sesgo negativo, donde los rendimientos extremos a la baja reducen el promedio, aunque la mayoría de los días presentan retornos positivos.

El segundo grupo muestra un perfil menos heterogéneo respecto a la media, con XOP registrando el peor desempeño promedio ($-0,0144\%$) —siendo el único negativo—, mientras que XLE alcanza el mayor ($0,0105\%$). Similar al grupo anterior, las medianas mantienen valores consistentemente positivos, lo que refuerza la noción de una tendencia subyacente alcista interrumpida ocasionalmente por caídas pronunciadas.

La evaluación de la desviación estándar revela una clara diferencia entre ambos grupos. El conjunto de energías limpias muestra volatilidades más dispersas, con los valores mínimo y máximo más extremos de toda la muestra (FAN = $1,6964\%$ y TAN = $2,8408\%$). En consecuencia, este grupo resulta más volátil en promedio, aunque el riesgo medio de las energías convencionales es ligeramente inferior.

El análisis de la asimetría confirma que todas las distribuciones presentan sesgo negativo, con valores que van desde $-0,2559$ (PBW) hasta $-1,2901$ (XOP). Esto implica una mayor probabilidad de observar rendimientos negativos extremos en comparación con los positivos, un aspecto crítico en la gestión del riesgo de cola.

La curtosis, significativamente superior a 3 en todos los casos, indica distribuciones leptocúrticas, caracterizadas por colas pesadas y picos agudos. El caso más extremo es XOP, con una curtosis de $28,6562$, lo que sugiere una frecuencia inusualmente alta de eventos extremos. Los estadísticos de Jarque-Bera, todos con p-valores de $0,0000$, permiten rechazar contundentemente la hipótesis de normalidad. Este ETF también destaca por registrar el valor mínimo más bajo, al

perder casi un 46 % de su valor en una sola jornada.

Los resultados de las pruebas de raíz unitaria (ADF y PP) confirman la estacionariedad de todas las series, con p-valores de 0,0000 en la prueba ADF y estadísticos altamente negativos en la de Phillips-Perron. En cuanto a la prueba KPSS, los valores del estadístico —que oscilan entre 0,0683 y 0,4061— se encuentran por debajo de los valores críticos al 5 % (0,463) y 1 % (0,739). Por tanto, no se rechaza la hipótesis nula de estacionariedad, lo que proporciona evidencia sólida de que las series de rendimientos no presentan tendencia estocástica. Este resultado respalda la idoneidad de los datos para el modelado de la volatilidad condicional mediante modelos de la familia GARCH.

El test de Ljung-Box muestra patrones diferenciados de autocorrelación. Mientras FAN e ICLN no evidencian autocorrelación significativa (p-valores de 0,1011 y 0,2648, respectivamente), los demás instrumentos exhiben dependencia temporal, especialmente IYE (p-valor 0,0000) y XOP (p-valor 0,0105). Esta autocorrelación sugiere posibles ineficiencias de mercado o la influencia persistente de factores estructurales en la dinámica de los precios.

	FAN	ICLN	IYE	PBW	QCLN
Media(%)	-0,0046	-0,0417	-0,0326	0,0069	-0,0244
Mediana(%)	0,0468	0,0000	0,0000	0,1063	0,0000
Máximo(%)	17,7464	19,7610	15,8199	14,8918	16,0005
Mínimo(%)	-13,5144	-20,7753	-15,6373	-15,4831	-16,7229
Dev.Std.(%)	1,6964	2,8408	2,3953	2,2316	2,0759
Asimetría	-0,4123	-0,2987	-0,2559	-0,2937	-0,5535
Curtosis	15,1478	8,8946	7,1266	7,5569	13,4525
Jarque-Bera (p)	0,0000	0,0000	0,0000	0,0000	0,0000
Ljung-Box Q (p)	0,1011	0,0408	0,0072	0,0190	0,2648
ADF (p)	0,0000	0,0000	0,0000	0,0000	0,0000
PP	-66,7690	-62,2789	-63,7174	-64,1210	-63,7677
KPSS	0,4061	0,3221	0,1863	0,1613	0,4060

Cuadro 1: Estadísticas descriptivas ETFs energías limpias.

La Figura ?? muestra todas las series de ETFs con transformación logarítmica porcentual:

	TAN	VDE	XLE	XOP
Media(%)	0,0078	0,0105	-0,0144	0,0072
Mediana(%)	0,0496	0,0575	0,0577	0,0469
Máximo(%)	15,7835	15,2503	19,8194	21,8133
Mínimo(%)	-22,0964	-22,4910	-45,9985	-23,1433
Dev.Std.(%)	1,9804	1,9581	2,6684	1,9609
Asimetría	-0,6904	-0,7371	-1,2901	-0,7589
Curtosis	15,3519	16,9457	28,6562	21,5574
Jarque-Bera (p)	0,0000	0,0000	0,0000	0,0000
Ljung-Box Q (p)	0,0275	0,0413	0,0105	0,0000
ADF (p)	0,0000	0,0000	0,0000	0,0000
PP	-68,5727	-68,6053	-67,8604	-70,4105
KPSS	0,0991	0,0953	0,0683	0,0903

Cuadro 2: Estadísticas descriptivas segunda mitad.

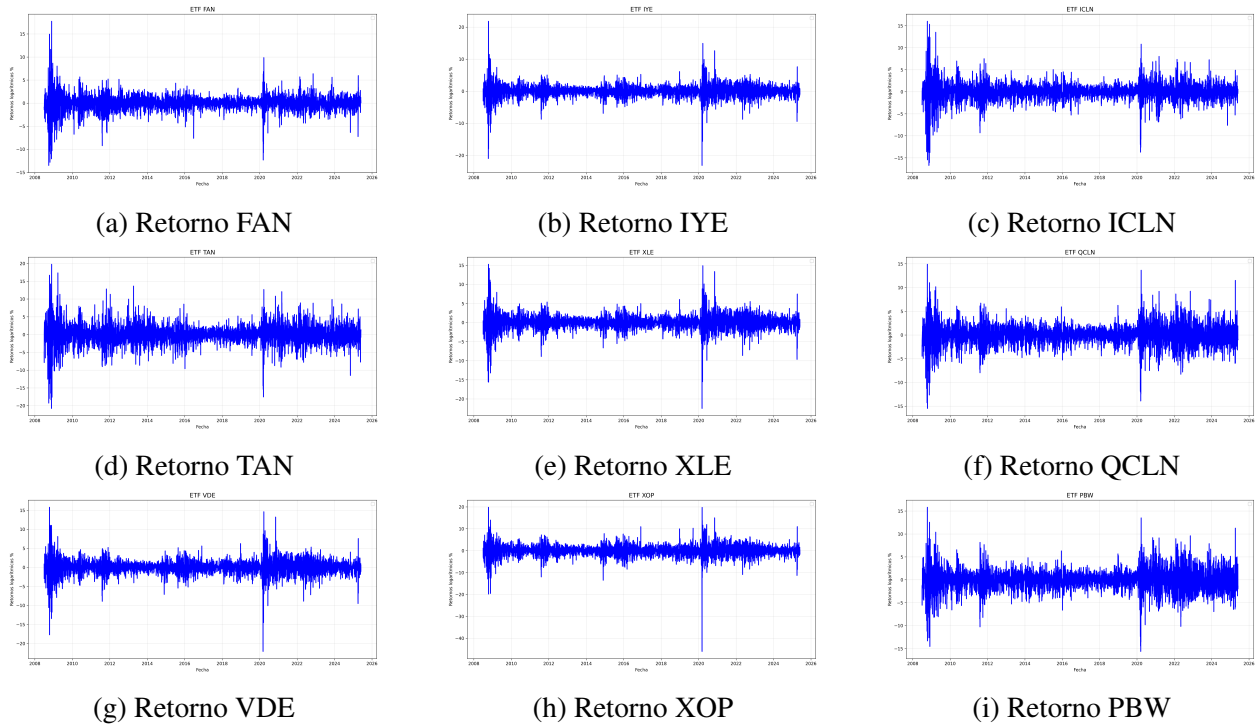


Figura 2: Retornos Logarítmicos %

5.2. FAN

5.2.1. Entrenamiento modelos familia GARCH FAN (σ^2)

Del proceso de optimización se obtuvieron los cinco modelos presentados en la Tabla 3. Entre ellos, el que alcanza el mejor desempeño según el criterio AIC es el EGARCH. Cabe destacar que los modelos que incluyen la innovación de orden o evidencian una capacidad explicativa del efecto apalancamiento de hasta cinco rezagos diarios.

En particular, el resultado a favor del EGARCH sugiere que la dinámica de la volatilidad en la serie analizada no es simétrica, es decir, los choques negativos y positivos no impactan de la misma forma sobre la varianza condicional.

Modelo	p	o	q	vix	AIC
GARCH	1	-	1	-	10942,555
GJR-GARCH	3	4	4	-	10931,059
EGARCH	1	5	1	-	10916,555
GARCH VIX	4	-	4	1	10996,553
GJR-GARCH VIX	4	4	5	1	10976,554

Cuadro 3: Especificación de rezagos óptimos para modelos de familia GARCH FAN.

Los indicadores de desempeño en el entrenamiento reflejan un ajuste limitado sobre la muestra, tal como señala la literatura. Dentro de los modelos clásicos, el GJR-GARCH presenta el mejor rendimiento con un R^2 de 24,37 %, un MSE de 117,68 y métricas ajustadas por heterocedasticidad de 7562,117 (HMSE) y 23,161 (HMAE). Si bien puede considerarse el más destacado en esta etapa, los modelos que incorporan la variable VIX muestran un mejor comportamiento en las métricas QLIKE y en la predicción de cambio de dirección.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	22,718	120,238	3,357	8036,117	23,506	1,630	0,285
GJR-GARCH	24,365	117,675	3,340	7562,117	23,161	1,626	0,302
EGARCH	24,011	118,227	3,252	7995,449	23,377	1,619	0,360
GARCH VIX	23,447	119,104	3,212	9476,284	24,046	1,611	0,401
GJR-GARCH VIX	23,926	118,359	3,209	9143,726	23,714	1,603	0,374

Cuadro 4: Métricas de desempeño en entrenamiento familia GARCH FAN (σ^2).

5.2.2. Entrenamiento modelos Híbridos Varianza FAN

Como se aprecia en la Tabla 5, los modelos tienden a utilizar pocos rezagos para explicar la dinámica de la volatilidad. Exceptuando el híbrido GJR-GARCH, todos descartan algún porcentaje de neuronas mediante regularización. En relación con el resto de los hiperparámetros, no se observa un patrón sistemático que permita establecer tendencias claras.

Modelo LSTM	Rezagos	Neuronas C1	Neuronas C2	Neuronas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	2	16	24	8	0,4	0,01	128	70
GJR-GARCH	2	24	30	12	0	0,01	128	30
EGARCH	2	16	20	8	0,4	0,001	64	70
GARCH VIX	4	4	12	2	0,2	0,00001	64	40
GJR-GARCH VIX	2	24	4	60	0,4	0,01	128	40

Cuadro 5: Hiperparámetros óptimos para varianza de FAN.

Los modelos híbridos de la Tabla 6 exhiben un desempeño deficiente en términos de bondad de ajuste. En particular, el LSTM-GARCH muestra la menor capacidad explicativa de todos los modelos, tanto clásicos como híbridos, durante el periodo de entrenamiento. En general, los híbridos obtienen peores resultados en la mayoría de las métricas, con la excepción del MDA, donde todos superan al mejor de los modelos clásicos.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	16,168	130,429	3,110	5593,369	23,612	1,717	0,441
LSTM GJR-GARCH	23,597	118,870	3,165	6550,108	23,089	1,642	0,444
LSTM EGARCH	20,258	124,066	3,122	5927,254	23,444	1,661	0,447
LSTM GARCH VIX	6,324	145,744	3,126	7850,102	22,864	1,774	0,496
LSTM GJR-GARCH VIX	21,612	121,959	3,145	10471,863	24,307	1,622	0,457

Cuadro 6: Métricas de desempeño en entrenamiento FAN (σ^2).

5.2.3. Comparación de modelos fuera de muestra para Varianza FAN

La Tabla 7 evidencia la escasa capacidad de los modelos para mantener su poder explicativo fuera de la muestra de entrenamiento. Incluso, el GARCH clásico arroja un R^2 negativo, lo que indica que su predicción resulta peor que asumir la media de la volatilidad diaria. No obstante,

a nivel de MSE, MAE, HMSE y HMAE, los modelos híbridos superan al mejor de los modelos clásicos, salvo el LSTM-GJR-GARCH-VIX, que es superado en HMSE y HMAE por el GJR-GARCH. En cuanto al MDA, los híbridos también logran mejores resultados, aunque no presentan superioridad en QLIKE.

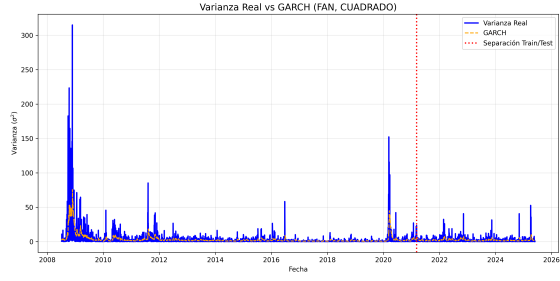
Modelo	$R^2 \%$	MSE $\% ^2$	MAE $\% ^2$	HMSE $\% ^2$	HMAE $\% ^2$	QLIKE	MDA
GARCH	-0,824	15,413	2,100	30.106,778	45,157	1,622	0,288
GJR-GARCH	0,547	15,203	2,080	26.266,876	43,929	1,616	0,304
EGARCH	1,491	15,059	2,084	30.225,355	46,011	1,599	0,363
GARCH-VIX	1,953	14,988	2,111	27.150,229	45,273	1,590	0,384
GJR-GARCH-VIX	2,565	14,895	2,096	30.618,631	46,210	1,585	0,388
LSTM-GARCH	2,888	14,845	1,964	17.351,096	39,771	1,589	0,452
LSTM-GJR-GARCH	3,302	14,782	2,037	21.051,763	42,209	1,608	0,456
LSTM-EGARCH	3,985	14,678	2,003	21.904,727	42,707	1,577	0,447
LSTM-GARCH-VIX	3,979	14,678	1,996	20.788,375	41,650	1,607	0,520
LSTM-GJR-GARCH-VIX	4,970	14,527	2,091	26.976,654	45,738	1,586	0,462

Cuadro 7: Resultados de métricas de desempeño FAN (σ^2).

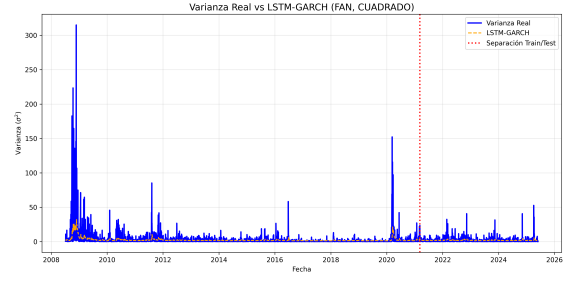
La aparente ventaja de los modelos híbridos debe interpretarse con cautela. La prueba de conjunto superior muestra que solo los modelos GJR-GARCH y GARCH resultan estadísticamente inferiores con un nivel de significancia del 5 %. En consecuencia, no es posible afirmar con certeza una dominancia de los modelos LSTM en términos de MSE.

Modelo	Posición	Estadístico	valor-p
LSTM-EGARCH	1	-1,34267	1
LSTM-GJR-GARCH-VIX	2	-1,25502	1
LSTM-GARCH-VIX	3	-0,77973	1
LSTM-GJR-GARCH	4	-0,18126	1
LSTM-GARCH	5	0,239764	0,9996
GJR-GARCH-VIX	6	0,81698	0,8676
GARCH-VIX	7	1,546373	0,361
EGARCH	8	1,907174	0,1864

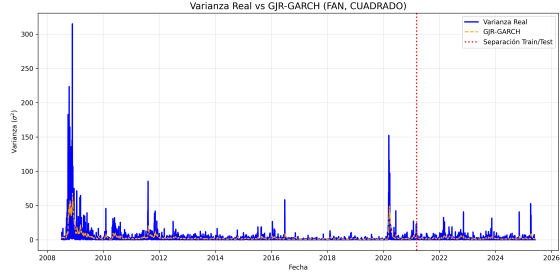
Cuadro 8: Conjunto superior de modelos para varianza de FAN.



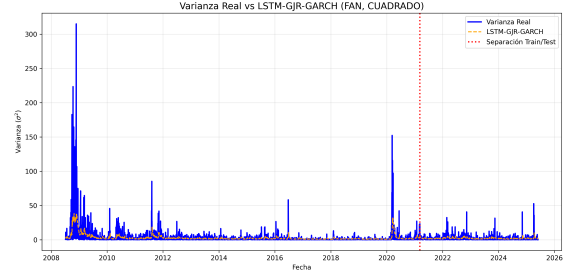
(a) Varianza - GARCH



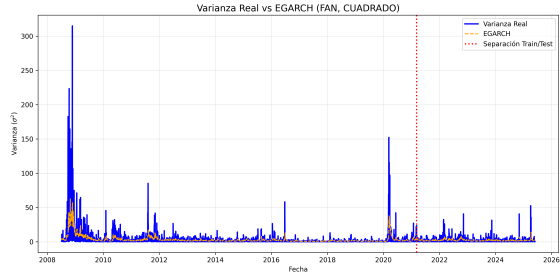
(b) Varianza - LSTM GARCH



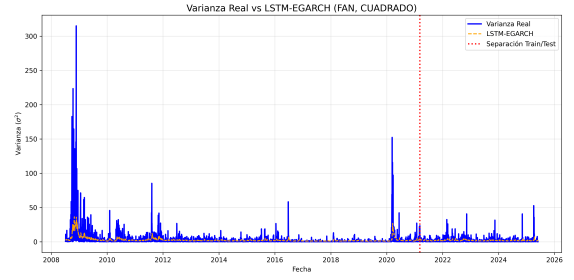
(c) Varianza - GJR-GARCH



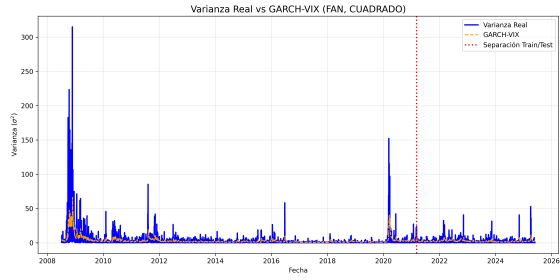
(d) Varianza - LSTM GJR-GARCH



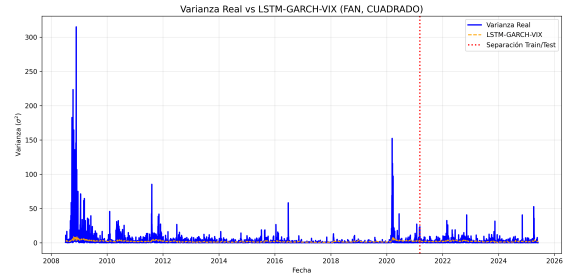
(e) Varianza - EGARCH



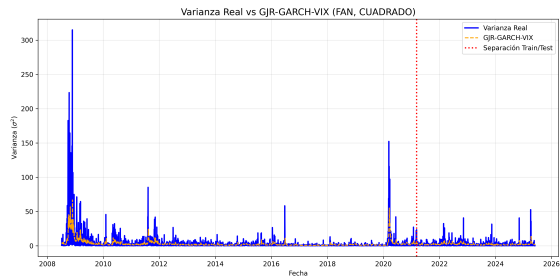
(f) Varianza - LSTM EGARCH



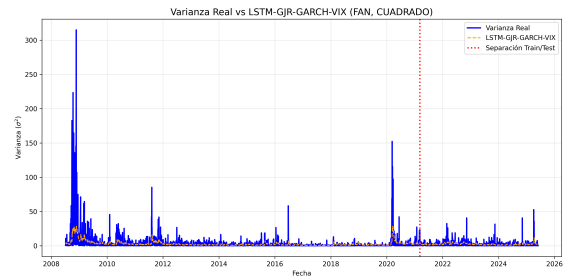
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 3: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF FAN (σ^2).

5.2.4. Entrenamiento modelos familia GARCH FAN (σ)

En esta etapa se aplica la raíz cuadrada sobre la serie, sin efectuar un nuevo proceso de reentrenamiento para los modelos de la familia GARCH. Aun así, los resultados indican que los modelos mantienen su capacidad de ajuste e incluso presentan leves mejoras en las métricas de desempeño, con excepción del GJR-GARCH, cuyo rendimiento disminuye marginalmente.

El modelo con mejor desempeño global dentro de la muestra de entrenamiento es el GARCH con la variable VIX, destacando en las métricas de R^2 %, MSE, MAE y MDA. Sin embargo, el modelo GJR-GARCH obtiene mejores resultados en la métrica HMSE, mientras que GJR-GARCH VIX se posiciona como el más eficiente en QLIKE, lo que sugiere una mejor representación del componente de error logarítmico.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	22,960	1,460	0,838	19,276	2,181	1,083	0,284
GJR-GARCH	23,673	1,446	0,835	18,953	2,169	1,081	0,303
EGARCH	26,033	1,401	0,826	19,178	2,165	1,079	0,359
GARCH VIX	27,118	1,381	0,820	19,849	2,165	1,077	0,401
GJR-GARCH VIX	26,695	1,389	0,816	19,542	2,152	1,075	0,375

Cuadro 9: Métricas de desempeño en entrenamiento familia GARCH FAN (σ).

5.2.5. Entrenamiento modelos híbridos Desviación Estándar FAN

En los modelos híbridos LSTM-GARCH, se observa que el modelo base GARCH requiere un número considerable de rezagos (13), lo que indica una estructura temporal más compleja. En contraste, el resto de los modelos presenta arquitecturas más simples, con uno o dos rezagos, y en su mayoría aplican regularización mediante descarte aleatorio (*dropout*). El modelo GARCH es el único que no incorpora esta técnica, lo que podría afectar su capacidad de generalización.

En cuanto a la tasa de aprendizaje, los modelos presentan valores relativamente altos (del orden de 10^{-2}). El modelo GARCH VIX destaca por ser el más complejo en términos de número de neuronas, lo que sugiere una mayor capacidad de aprendizaje no lineal y adaptación a la información exógena incorporada.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Tasa Aprendizaje	Lote	Épocas
GARCH	13	20	12	8	0	0,01	64	50
GJR-GARCH	2	16	8	4	0,4	0,01	64	40
EGARCH	2	4	24	12	0,4	0,01	128	60
GARCH VIX	1	24	30	4	0,2	0,001	64	50
GJR-GARCH VIX	2	12	8	8	0,2	0,01	64	70

Cuadro 10: Hiperparámetros óptimos para desviación estándar de FAN.

Los resultados presentados en la Tabla 11 muestran que los modelos híbridos superan de forma consistente a sus contrapartes clásicas en todas las métricas durante el periodo de entrenamiento. Esto sugiere que la combinación de estructuras GARCH con redes LSTM mejora la capacidad del modelo para capturar dependencias no lineales y patrones de volatilidad más complejos.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	24,945	1,422	0,731	10,134	1,607	1,068	0,385
LSTM GJR-GARCH	30,820	1,311	0,732	11,384	1,700	1,062	0,424
LSTM EGARCH	32,583	1,277	0,721	10,717	1,635	1,055	0,411
LSTM GARCH VIX	30,824	1,311	0,713	10,728	1,572	1,054	0,446
LSTM GJR-GARCH VIX	30,719	1,313	0,716	10,229	1,588	1,062	0,445

Cuadro 11: Métricas de desempeño en entrenamiento modelos híbridos FAN (σ).

5.2.6. Comparación de modelos fuera de muestra para Desviación Estándar FAN

Al evaluar el desempeño fuera de muestra, se observa que ninguno de los diez modelos logra preservar completamente su capacidad predictiva. Si bien las métricas MSE y MAE disminuyen, este comportamiento se atribuye más a una menor volatilidad o a una menor dispersión de los datos que a una mejora real en la predicción. Esto queda en evidencia al analizar las métricas HMSE y HMAE, que empeoran para todos los modelos al compararse con los resultados dentro de la muestra.

Pese a lo anterior, los modelos híbridos superan sistemáticamente a los clásicos en las métricas QLIKE y MDA, demostrando una mejor capacidad para anticipar los cambios de dirección en la volatilidad. El modelo LSTM-GJR-GARCH-VIX destaca como el más robusto, dominando en prácticamente todos los indicadores de desempeño y mostrando la mayor estabilidad predictiva

fuera de muestra.

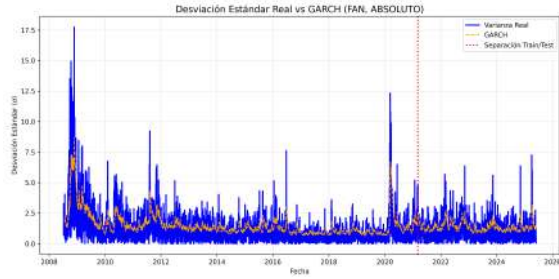
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-19,455	0,999	0,804	39,490	2,873	1,050	0,288
GJR-GARCH	-17,472	0,983	0,799	38,313	2,847	1,048	0,304
EGARCH	-17,366	0,982	0,800	40,292	2,897	1,048	0,363
GARCH-VIX	-18,848	0,994	0,810	39,527	2,908	1,051	0,384
GJR-GARCH-VIX	-17,718	0,985	0,804	40,469	2,906	1,048	0,388
LSTM-GARCH	2,542	0,815	0,672	20,195	2,059	1,008	0,390
LSTM-GJR-GARCH	2,962	0,812	0,681	21,584	2,140	1,007	0,434
LSTM-EGARCH	3,531	0,807	0,678	21,620	2,124	1,007	0,413
LSTM-GARCH-VIX	5,026	0,795	0,669	20,697	2,062	1,002	0,452
LSTM-GJR-GARCH-VIX	5,805	0,788	0,659	18,729	1,988	0,999	0,456

Cuadro 12: Resultados de métricas de desempeño FAN (σ).

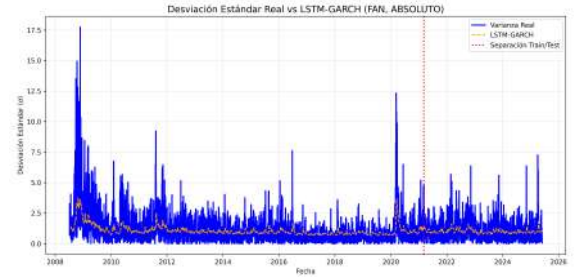
Finalmente, la superioridad de los modelos híbridos queda respaldada por la prueba de conjunto superior (MCS), la cual elimina a todos los modelos clásicos y al LSTM-GJR-GARCH, confirmando que los híbridos dominan estadísticamente en términos de MSE con un nivel de significancia del 5 %. Este resultado valida la efectividad de las arquitecturas híbridas para modelar dinámicas complejas de volatilidad en la serie analizada.

Modelo	Posición	Estadístico	valor-p
LSTM-GJR-GARCH-VIX	1	-3,17525	1
LSTM-GARCH-VIX	2	-1,03960	1
LSTM-EGARCH	3	1,31954	0,3638
LSTM-GARCH	4	2,18308	0,0620

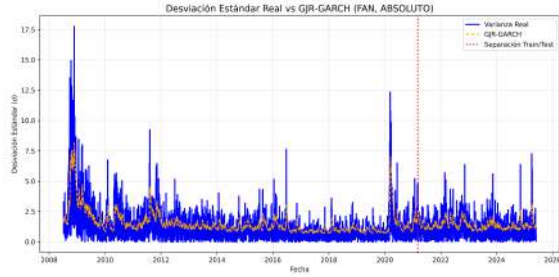
Cuadro 13: Conjunto superior de modelos para desviación estándar de FAN.



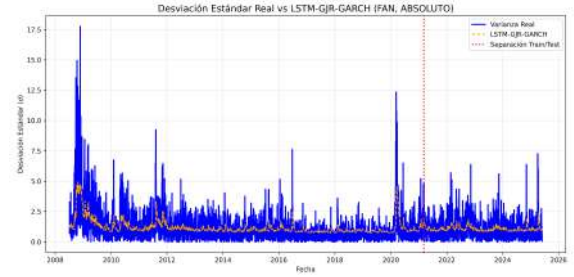
(a) Desviación estándar - GARCH



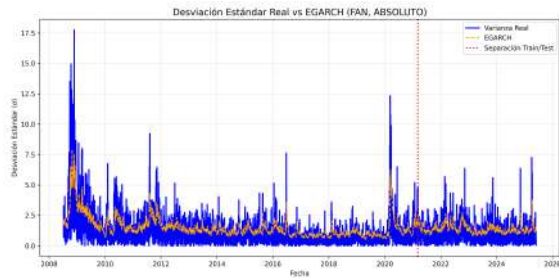
(b) Desviación estándar - LSTM GARCH



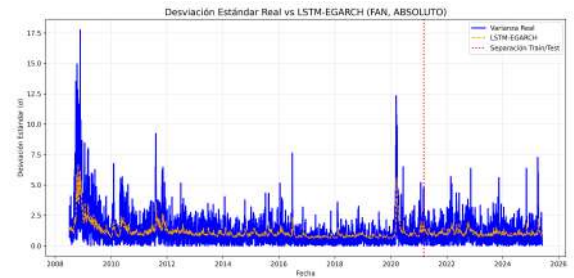
(c) Desviación estándar - GJR-GARCH



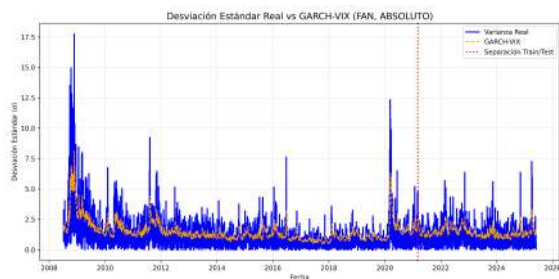
(d) Desviación estándar - LSTM GJR-GARCH



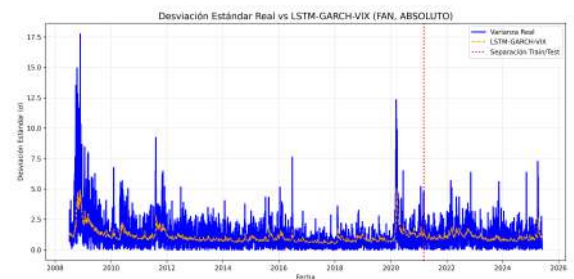
(e) Desviación estándar - EGARCH



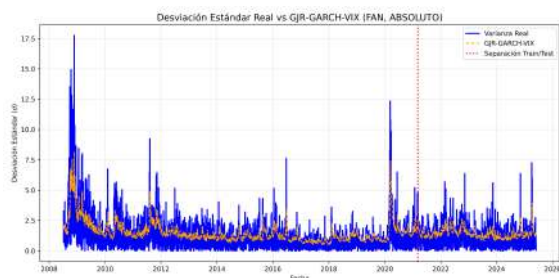
(f) Desviación estándar - LSTM EGARCH



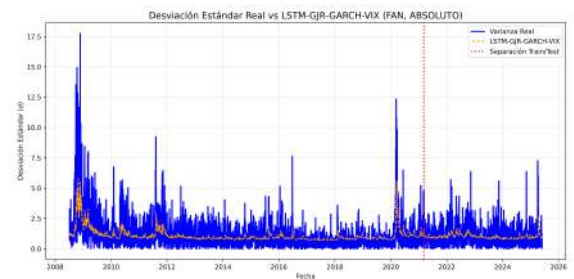
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.3. ICLN

5.3.1. Entrenamiento modelos familia GARCH ICLN (σ^2)

Del proceso de optimización de la matriz de rezagos se obtuvieron los modelos con mejor ajuste según el criterio AIC, presentados en la Tabla 5. El modelo con el menor valor de AIC corresponde al GJR-GARCH, de estructura simétrica y orden (2,2,2), lo que evidencia una capacidad superior para capturar los efectos de volatilidad condicional. Por su parte, los modelos que incorporan la variable exógena VIX presentan un rezago en dicha variable, por lo que no requieren de múltiples de esta para explicar la volatilidad.

Modelo	p	o	q	VIX	AIC
GARCH	1	-	1	-	12085,201
GJR-GARCH	2	2	2	-	12079,377
EGARCH	3	3	4	-	12120,173
GARCH VIX	4	-	3	1	12115,049
GJR-GARCH VIX	4	4	2	1	12098,899

Cuadro 14: Especificación de rezagos óptimos para modelos de familia GARCH ICLN.

En general, los modelos presentan una bondad de ajuste superior al 25 %, aunque con valores de MDA inferiores a 0,4, lo que indica una limitada capacidad de predicción direccional. El GJR-GARCH con la variable VIX muestra el mejor desempeño en las métricas especializadas en volatilidad (HMSE, HMAE y QLIKE), mientras que el GJR-GARCH sin VIX obtiene los mejores resultados en $R^2\%$ y MSE, y el EGARCH se destaca marginalmente en MAE. Estos resultados sugieren que la inclusión de la variable VIX mejora la modelación de la heterocedasticidad, aunque no necesariamente incrementa la precisión explicativa global.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	28,573	219,134	4,985	73893,016	32,626	1,997	0,293
GJR-GARCH	28,854	218,271	4,979	70893,821	32,540	1,992	0,383
EGARCH	25,490	228,593	4,972	62268,325	32,291	1,999	0,369
GARCH VIX	26,502	225,487	4,794	66139,487	32,043	1,979	0,360
GJR-GARCH VIX	27,053	223,797	4,808	60764,677	31,711	1,975	0,397

Cuadro 15: Métricas de desempeño en entrenamiento familia GARCH ICLN (σ^2).

5.3.2. Entrenamiento modelos híbridos Varianza ICLN

En el caso de los modelos híbridos, la serie ICLN presenta una marcada preferencia por arquitecturas con múltiples rezagos y un alto número de neuronas en las capas ocultas, especialmente en el modelo GJR-GARCH VIX, que utiliza 9 rezagos y hasta 50 neuronas por capa. En términos de regularización, la mayoría de los modelos muestra una baja o nula aplicación de descarte aleatorio (*dropout*), con excepción del modelo GJR-GARCH, que utiliza una proporción moderada (0,4). Por otro lado, las tasas de aprendizaje son elevadas (en torno a 10^{-2}), lo que sugiere una estrategia de entrenamiento más agresiva orientada a una convergencia rápida.

Modelo LSTM	Rezagos	Neuro- nas C1	Neuro- nas C2	Neuro- nas C3	Descarte Aleatorio	Tasa Aprendizaje	Lote	Épocas
GARCH	7	16	4	16	0,2	0,01	64	60
GJR-GARCH	3	24	16	20	0,4	0,001	128	70
EGARCH	12	12	4	16	0,2	0,01	128	30
GARCH VIX	1	4	40	2	0	0,01	64	60
GJR-GARCH VIX	9	4	50	50	0	0,01	64	60

Cuadro 16: Hiperparámetros óptimos para varianza de ICLN.

Los resultados del entrenamiento muestran una mejora moderada de los modelos híbridos en comparación con los clásicos, particularmente en las métricas R^2 %, MSE y MAE. El incremento más significativo se observa en las métricas ajustadas por heterocedasticidad (HMSE y HMAE), donde los híbridos presentan un rendimiento considerablemente superior, lo que evidencia una mejor captura de la variabilidad condicional. Asimismo, los modelos LSTM exhiben un mejor desempeño promedio en MDA, aunque presentan valores levemente peores en QLIKE. Si bien no se identifica un modelo que domine en todos los indicadores, el LSTM-GARCH destaca al liderar en tres de las seis métricas evaluadas.

5.3.3. Comparación de modelos fuera de muestra para Varianza ICLN

Los resultados de la Tabla 18 muestran que los modelos híbridos mantienen un mejor desempeño global fuera de muestra en comparación con los clásicos. Todos los híbridos superan al mejor modelo tradicional en R^2 %, MSE, HMSE y MDA, lo que confirma una mayor capacidad de

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	29,659	215,803	4,615	43905,261	28,206	2,018	0,384
LSTM GJR-GARCH	26,399	225,802	5,090	46223,710	36,131	2,062	0,461
LSTM EGARCH	32,313	207,660	4,700	54791,698	29,172	2,018	0,420
LSTM GARCH VIX	26,842	224,443	4,903	51221,845	35,006	1,992	0,465
LSTM GJR-GARCH VIX	30,361	213,647	4,838	51922,120	32,494	1,985	0,445

Cuadro 17: Métricas de desempeño en entrenamiento modelos híbridos ICLN (σ^2).

generalización. En la métrica HMAE, los híbridos presentan tanto el mejor valor (40,263, LSTM-GARCH) como el peor (49,022, LSTM-GJR-GARCH), lo que evidencia cierta variabilidad entre arquitecturas. En el caso de QLIKE, las diferencias entre modelos son mínimas respecto al entrenamiento, indicando estabilidad en el componente de error logarítmico. De forma general, el modelo LSTM-GJR-GARCH VIX logra destacar en tres de las seis métricas, mostrando un equilibrio favorable entre precisión y robustez.

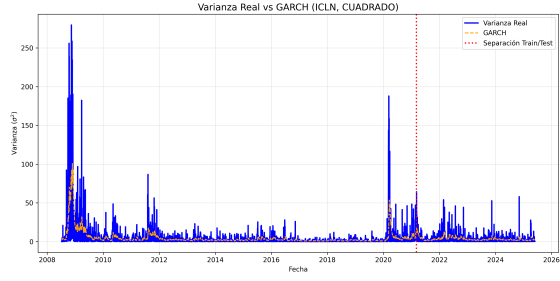
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	3,509	36,658	3,375	123770,163	46,186	2,104	0,294
GJR-GARCH	4,642	36,228	3,335	136139,023	46,876	2,095	0,389
EGARCH	4,298	36,358	3,445	109613,437	47,963	2,106	0,362
GARCH VIX	4,788	36,173	3,336	87940,330	44,439	2,092	0,339
GJR-GARCH VIX	5,077	36,063	3,318	101124,566	45,510	2,093	0,380
LSTM GARCH	5,928	35,739	3,139	81062,078	40,263	2,103	0,393
LSTM GJR-GARCH	6,030	35,701	3,539	80341,150	49,022	2,091	0,494
LSTM EGARCH	6,099	35,674	3,172	102544,430	41,941	2,110	0,423
LSTM GARCH VIX	5,800	35,788	3,371	48709,205	42,745	2,078	0,475
LSTM GJR-GARCH VIX	7,391	35,184	3,322	79332,683	44,558	2,077	0,445

Cuadro 18: Resultados de métricas de desempeño ICLN (σ^2).

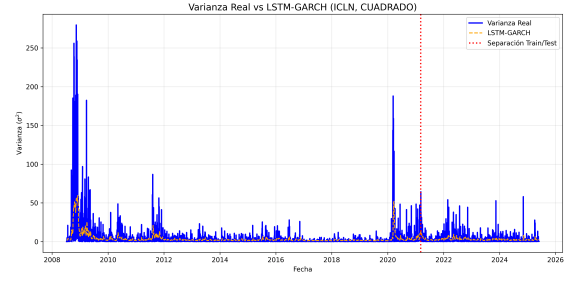
A pesar de que los modelos LSTM presentan las mejores métricas en el periodo de evaluación, las diferencias no son lo suficientemente significativas como para eliminar a todos los modelos clásicos según la prueba de conjunto superior (MCS). Solo el modelo GARCH resulta estadísticamente inferior al 5% de significancia, mientras que el resto de las especificaciones, tanto clásicas como híbridas, no muestran diferencias estadísticamente relevantes en términos de MSE. Esto sugiere que, si bien los híbridos logran un mejor ajuste, su ventaja no es concluyente a nivel inferencial.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-2,46909	1
LSTM EGARCH	2	-0,54815	1
LSTM GJR-GARCH	3	-0,49891	1
LSTM GARCH	4	-0,4863	1
LSTM GARCH VIX	5	-0,11725	1
GJR-GARCH VIX	6	0,886097	0,9088
GARCH VIX	7	1,116728	0,7684
GJR-GARCH	8	1,261282	0,6518
EGARCH	9	1,583894	0,3916

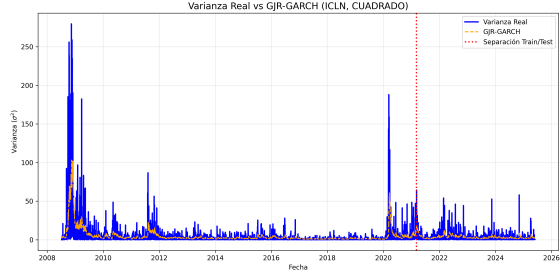
Cuadro 19: Conjunto superior de modelos para varianza de ICLN.



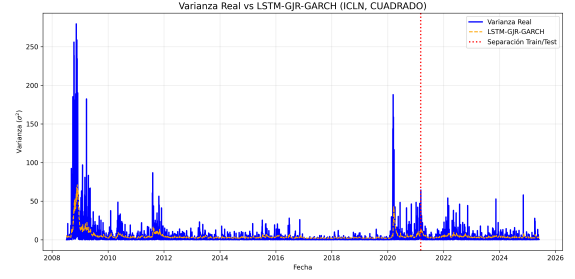
(a) Varianza - GARCH



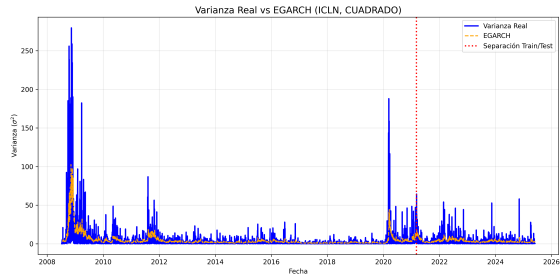
(b) Varianza - LSTM GARCH



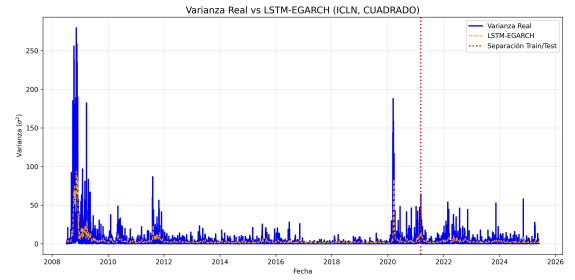
(c) Varianza - GJR-GARCH



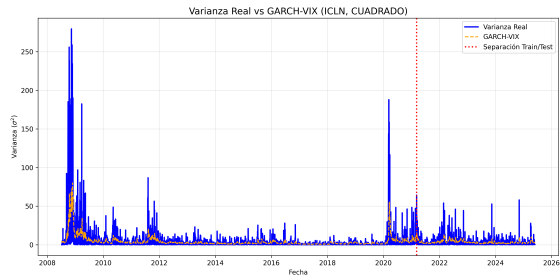
(d) Varianza - LSTM GJR-GARCH



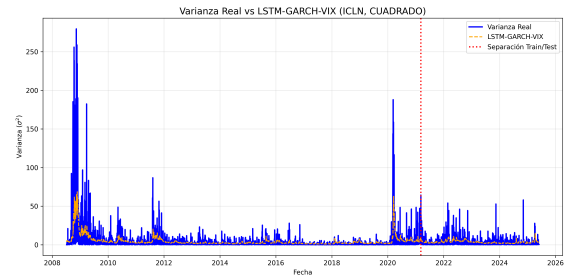
(e) Varianza - EGARCH



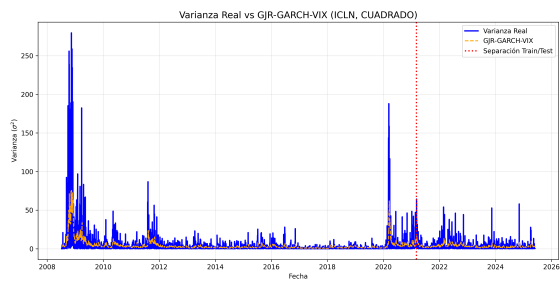
(f) Varianza - LSTM EGARCH



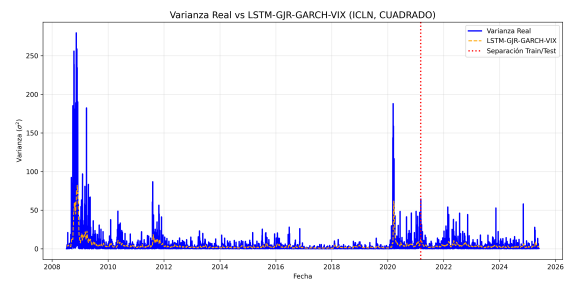
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 5: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF ICLN (σ^2).

5.3.4. Entrenamiento modelos familia GARCH ICLN (σ)

El proceso de entrenamiento para los modelos de la familia GARCH aplicados a la desviación estándar de la serie ICLN muestra un comportamiento estable en las métricas de desempeño, con valores de $R^2\%$ cercanos al 25% y errores promedio relativamente bajos. Tal como se observa en la Tabla 20, el mejor desempeño global corresponde a los modelos que incorporan la variable exógena VIX, particularmente el GJR-GARCH VIX, que logra los menores valores de MSE, MAE y QLIKE, así como el mejor ajuste en las métricas heterocedásticas (HMSE y HMAE).

En comparación, los modelos sin VIX mantienen un rendimiento competitivo, aunque inferior, siendo el GJR-GARCH el más destacado dentro de este grupo por su capacidad de modelar asimetrías leves en la respuesta de la volatilidad frente a choques negativos. En conjunto, todos los modelos presentan un desempeño homogéneo, lo que refleja que la dinámica de la serie es relativamente predecible dentro del periodo de entrenamiento.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	24,342	2,118	1,017	28,273	2,202	1,269	0,293
GJR-GARCH	24,535	2,112	1,012	28,207	2,192	1,267	0,383
EGARCH	24,342	2,118	1,019	27,934	2,208	1,270	0,369
GARCH VIX	28,159	2,011	0,997	27,750	2,175	1,264	0,359
GJR-GARCH VIX	28,169	2,011	0,996	27,435	2,167	1,263	0,397

Cuadro 20: Métricas de desempeño en entrenamiento familia GARCH ICLN (σ).

5.3.5. Entrenamiento modelos híbridos Desviación Estándar ICLN

En los modelos híbridos LSTM-GARCH aplicados a la desviación estándar, se observa una clara preferencia por arquitecturas de complejidad intermedia, con un número de rezagos entre 3 y 7 y un uso extensivo de neuronas en las capas ocultas, alcanzando hasta 50 en el caso del modelo GJR-GARCH VIX. El descarte aleatorio (*dropout*) se mantiene bajo (0,2) en casi todos los modelos, lo que favorece la retención de información temporal relevante. Asimismo, todos los modelos comparten una tasa de aprendizaje relativamente alta (10^{-2}), lo que permite una convergencia rápida durante el proceso de optimización, sin evidenciar sobreajuste.

De acuerdo con la Tabla 22, los modelos híbridos mejoran consistentemente el ajuste respecto

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Tasa Aprendizaje	Lote	Épocas
GARCH	3	16	12	16	0	0,01	128	50
GJR-GARCH	7	30	24	2	0,2	0,01	64	70
EGARCH	5	30	16	16	0,2	0,01	128	50
GARCH VIX	1	16	12	24	0,2	0,01	64	60
GJR-GARCH VIX	3	50	16	16	0,2	0,01	64	40

Cuadro 21: Hiperparámetros óptimos para desviación estándar de ICLN.

de los modelos clásicos, alcanzando incrementos notables en $R^2\%$ (superiores al 30 %) y reducciones en MSE y MAE. Las métricas ajustadas por heterocedasticidad (HMSE y HMAE) también muestran mejoras sustanciales, lo que sugiere una mayor capacidad para modelar la varianza condicional. El LSTM GJR-GARCH VIX es el modelo con mejor desempeño global, liderando en las métricas de MSE, MAE y QLIKE, mientras que el LSTM GARCH VIX presenta un equilibrio sólido entre ajuste y estabilidad.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	31,075	1,929	0,891	16,230	1,666	1,244	0,433
LSTM GJR-GARCH	30,736	1,939	0,877	13,599	1,590	1,248	0,413
LSTM EGARCH	31,942	1,905	0,895	15,638	1,699	1,252	0,389
LSTM GARCH VIX	32,505	1,889	0,883	16,889	1,691	1,239	0,426
LSTM GJR-GARCH VIX	34,055	1,846	0,882	17,327	1,659	1,238	0,411

Cuadro 22: Métricas de desempeño en entrenamiento modelos híbridos ICLN (σ).

5.3.6. Comparación de modelos fuera de muestra para Desviación Estándar ICLN

Los resultados fuera de muestra, presentados en la Tabla 23, confirman la superioridad de los modelos híbridos respecto a los clásicos en prácticamente todas las métricas. Mientras los modelos GARCH tradicionales muestran pérdidas significativas de capacidad predictiva (incluso con valores negativos de $R^2\%$), los híbridos logran mantener un desempeño estable, con mejoras consistentes en MSE, MAE y MDA. Las métricas heterocedásticas (HMSE y HMAE) también reflejan una reducción sustancial del error, indicando una mejor adaptación a los episodios de alta volatilidad fuera del periodo de entrenamiento.

Dentro del conjunto híbrido, el LSTM GARCH VIX y el LSTM GJR-GARCH VIX son

los modelos más destacados, con los menores valores de MSE y QLIKE y una mayor precisión direccional (MDA cercana a 0,44). El resto de los modelos híbridos conserva un rendimiento competitivo, lo que sugiere una estabilidad estructural en la arquitectura propuesta.

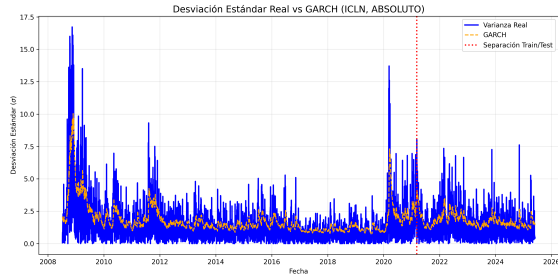
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-12,083	1,536	0,991	41,101	2,581	1,303	0,294
GJR-GARCH	-10,290	1,511	0,983	41,800	2,576	1,300	0,389
EGARCH	-14,169	1,565	1,013	42,698	2,668	1,309	0,362
GARCH VIX	-10,306	1,512	0,986	39,369	2,573	1,302	0,339
GJR-GARCH VIX	-9,653	1,503	0,982	40,437	2,575	1,301	0,380
LSTM GARCH	5,364	1,297	0,849	22,806	1,929	1,271	0,464
LSTM GJR-GARCH	5,563	1,294	0,834	20,362	1,833	1,270	0,421
LSTM EGARCH	5,892	1,290	0,839	22,484	1,898	1,268	0,393
LSTM GARCH VIX	6,991	1,275	0,842	20,652	1,880	1,266	0,439
LSTM GJR-GARCH VIX	6,918	1,276	0,838	21,833	1,876	1,268	0,407

Cuadro 23: Resultados de métricas de desempeño ICLN (σ).

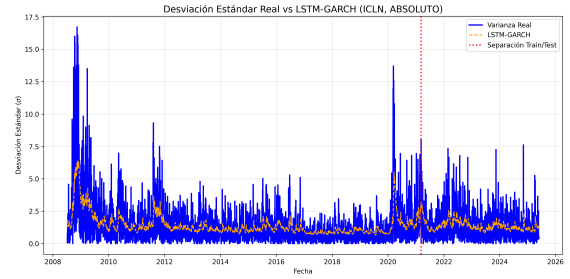
Finalmente, la prueba de conjunto superior (MCS) ratifica la dominancia estadística de los modelos híbridos sobre los clásicos, eliminándolos de este. En particular, los modelos LSTM GJR-GARCH VIX y LSTM GARCH VIX ocupan las primeras posiciones, confirmando su superioridad a un nivel de significancia del 5 %.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,38432	1
LSTM GARCH VIX	2	-0,92024	1
LSTM EGARCH	3	0,441409	0,9636
LSTM GJR-GARCH	4	1,010851	0,629
LSTM GARCH	5	1,318327	0,4202

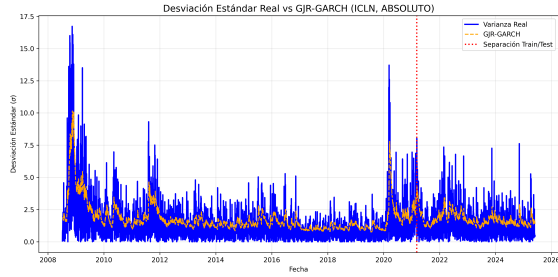
Cuadro 24: Conjunto superior de modelos para desviación estándar de ICLN.



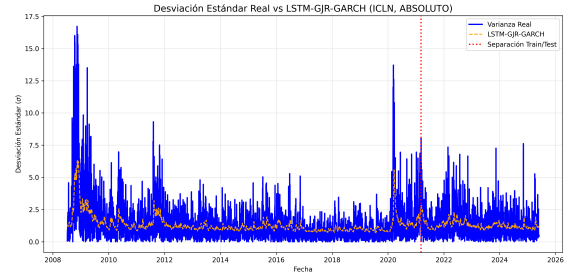
(a) Desviación estándar - GARCH



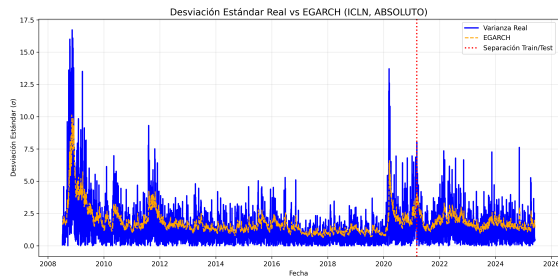
(b) Desviación estándar - LSTM GARCH



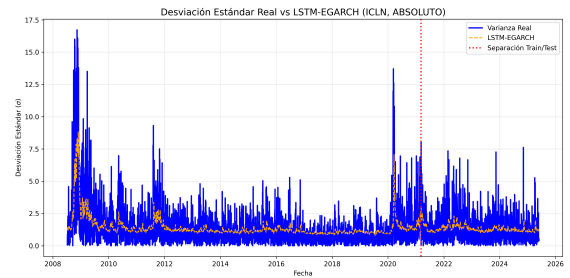
(c) Desviación estándar - GJR-GARCH



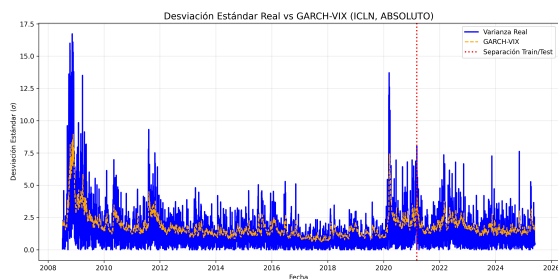
(d) Desviación estándar - LSTM GJR-GARCH



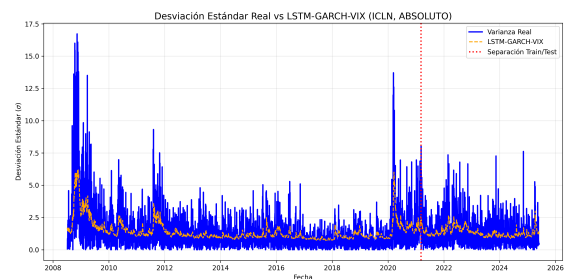
(e) Desviación estándar - EGARCH



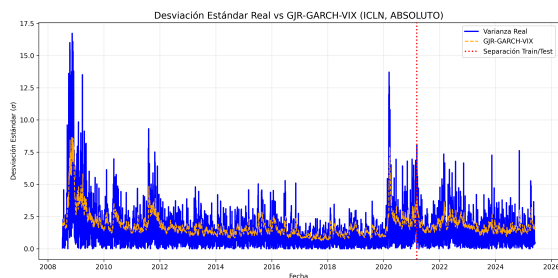
(f) Desviación estándar - LSTM EGARCH



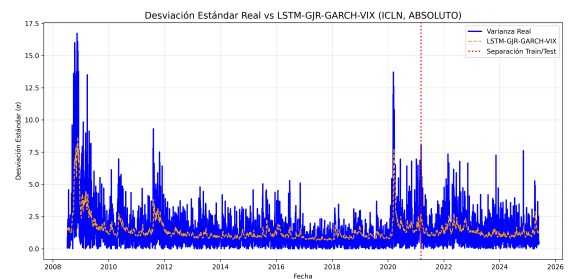
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.4. IYE

5.4.1. Entrenamiento modelos familia GARCH IYE (σ^2)

El proceso de optimización de la matriz de rezagos para la serie IYE muestra que los modelos GARCH sin la variable exógena VIX obtienen mejores resultados en términos del criterio de información de Akaike (AIC). En general, las especificaciones óptimas presentan órdenes en torno a tres rezagos en los términos autorregresivos y de medias móviles de la varianza, lo que evidencia una dinámica de volatilidad con cierta persistencia temporal, pero sin requerir estructuras excesivamente complejas.

De acuerdo con la Tabla 25, el modelo con mejor desempeño según el AIC es el EGARCH(3,2,5), seguido muy de cerca por el GJR-GARCH. La ausencia de mejoras al incorporar la variable VIX sugiere que la volatilidad implícita del mercado no tiene un impacto significativo sobre la varianza condicional de los retornos de IYE en el periodo analizado.

Modelo	p	o	q	vix	AIC
GARCH	1	-	1	-	11636,240
GJR-GARCH	3	3	1	-	11570,312
EGARCH	3	2	5	-	11567,109
GARCH VIX	3	-	3	1	11743,385
GJR-GARCH VIX	3	3	1	5	11867,477

Cuadro 25: Especificación de rezagos óptimos para modelos de familia GARCH IYE.

En términos de desempeño, los modelos presentan una bondad de ajuste relativamente baja, con valores de R^2 % cercanos inferiores al 20 %. El EGARCH sobresale con el mejor ajuste global, alcanzando un 19,07 % en R^2 % y los menores valores de MSE, MAE y QLIKE, además de liderar en cinco de las seis métricas evaluadas, excepto en HMSE. Este resultado confirma la capacidad del EGARCH para modelar asimetrías en la respuesta de la volatilidad ante choques negativos, reflejando el conocido efecto apalancamiento. Por su parte, los modelos que incorporan la variable VIX no logran mejoras sustanciales, lo que sugiere una débil relación contemporánea entre la volatilidad implícita del mercado y los movimientos de la varianza condicional en el ETF IYE.

Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	16,231	331,890	4,433	305425,059	89,033	1,812	0,279
GJR-GARCH	18,801	321,708	4,380	304421,883	87,090	1,791	0,404
EGARCH	19,070	320,643	4,173	262517,739	86,384	1,785	0,424
GARCH VIX	17,573	326,573	4,292	254571,214	87,242	1,806	0,318
GJR-GARCH VIX	17,701	326,065	4,410	324619,713	88,112	1,788	0,393

Cuadro 26: Métricas de desempeño en entrenamiento familia GARCH IYE (σ^2).

5.4.2. Entrenamiento modelos híbridos Varianza IYE

La arquitectura de los modelos híbridos LSTM-GARCH para la serie IYE se caracteriza por configuraciones moderadamente complejas, con un número de rezagos entre 3 y 4 y una cantidad considerable de neuronas en las capas ocultas (hasta 50 en el modelo GARCH-VIX). La regularización mediante descarte aleatorio (*dropout*) es utilizada en la mayoría de los casos con valores entre 0,2 y 0,4, lo que contribuye a evitar el sobreajuste en series con alta autocorrelación. En cuanto a la tasa de aprendizaje, predominan valores de 10^{-2} , salvo en dos configuraciones más conservadoras (10^{-4} y 10^{-3}), que buscan una convergencia más gradual en la función de pérdida.

Modelo LSTM	Rezagos	Neuronas C1	Neuronas C2	Neuronas C3	Descarte Aleatorio	Tasa Aprendizaje	Lote	Épocas
GARCH	4	30	20	4	0,4	0,01	128	70
GJR-GARCH	3	16	12	4	0,2	0,01	64	60
EGARCH	3	24	16	24	0,4	0,001	64	60
GARCH VIX	1	24	20	50	0,4	0,01	128	70
GJR-GARCH VIX	3	30	24	16	0,2	0,001	64	40

Cuadro 27: Hiperparámetros óptimos para varianza de IYE.

Los resultados de entrenamiento, mostrados en la Tabla 28, indican mejoras modestas en las métricas $R^2 \%$, MSE y MAE respecto de los modelos clásicos, aunque las ganancias más evidentes se observan en las métricas heterocedásticas (HMSE y HMAE), donde los híbridos reducen significativamente los errores de predicción en comparación con sus contrapartes tradicionales. El LSTM-EGARCH destaca como el modelo de mejor desempeño general, mostrando menores valores de MSE, MAE y QLIKE, mientras que el LSTM-GJR-GARCH-VIX exhibe una mejora en la predicción direccional (MDA) alcanzando valores cercanos a 0,48. Estos resultados sugieren que la combinación de componentes no lineales (LSTM) con estructuras de varianza condicional

logarítmica (EGARCH) mejora la sensibilidad ante cambios abruptos en la volatilidad.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	14,105	340,312	4,045	220795,178	94,133	1,886	0,449
LSTM GJR-GARCH	14,781	337,635	4,228	284070,198	107,904	1,922	0,479
LSTM EGARCH	17,038	328,693	4,037	182914,120	80,736	1,825	0,479
LSTM GARCH VIX	15,557	334,558	4,290	329129,409	101,407	1,852	0,473
LSTM GJR-GARCH VIX	16,418	331,150	4,189	265046,369	84,728	1,829	0,483

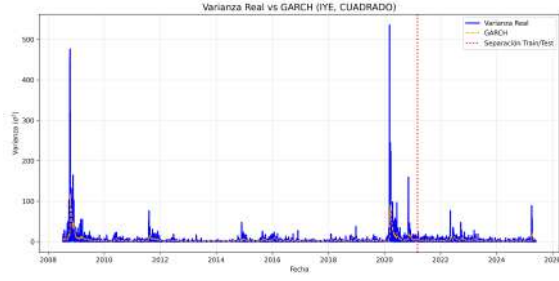
Cuadro 28: Métricas de desempeño en entrenamiento modelos híbridos IYE (σ^2).

5.4.3. Comparación de modelos fuera de muestra para Varianza IYE

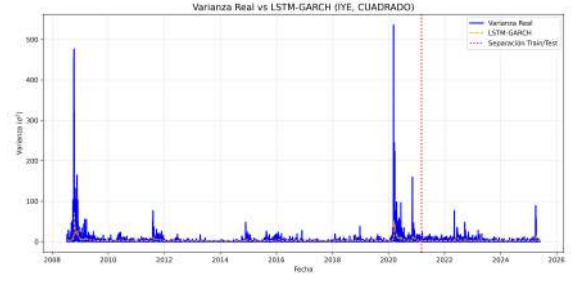
Los resultados fuera de muestra, presentados en la Tabla 29, confirman una mejora general de los modelos híbridos frente a los clásicos. Todos los modelos LSTM presentan valores superiores en R^2 %, MSE y MDA, lo que indica una mayor capacidad predictiva y estabilidad en entornos no observados. Sin embargo, la familia híbrida también exhibe la mayor dispersión en las métricas HMAE y QLIKE, reflejando que los modelos con mayor complejidad tienden a sobreajustar ciertas fluctuaciones locales.

Entre los modelos clásicos, el EGARCH continúa siendo el más robusto, dominando cinco de las seis métricas dentro de su familia y superando a varios híbridos en HMSE y HMAE. Aun así, los modelos LSTM-GARCH y LSTM-EGARCH mantienen las mejores combinaciones de bajo error y buena precisión direccional. El LSTM GJR-GARCH VIX es el que alcanza la mayor MDA (0,499) y un equilibrio sólido entre las métricas, por lo que puede considerarse el mejor compromiso entre ajuste y capacidad de generalización.

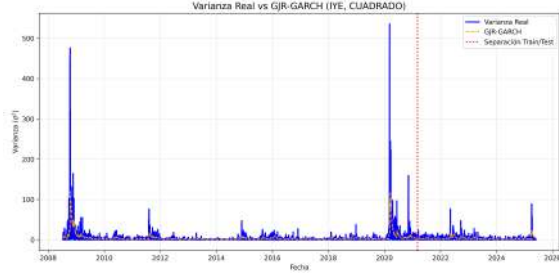
Finalmente, la prueba de conjunto superior (MCS) confirma que las diferencias entre modelos no son estadísticamente significativas al 5 % de confianza. Aunque los híbridos presentan un mejor rendimiento promedio, la mejora no es suficiente para eliminar a los modelos clásicos. En consecuencia, se concluye que los modelos LSTM aportan un valor agregado principalmente en estabilidad direccional (MDA), más que en reducción de error promedio.



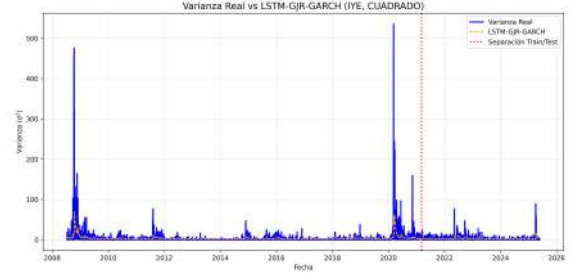
(a) Varianza - GARCH



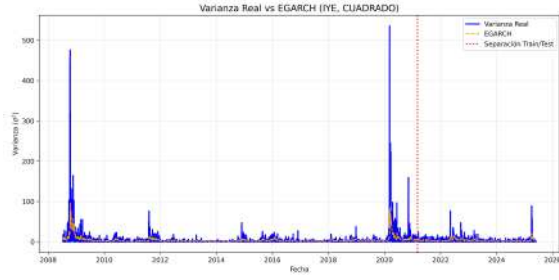
(b) Varianza - LSTM GARCH



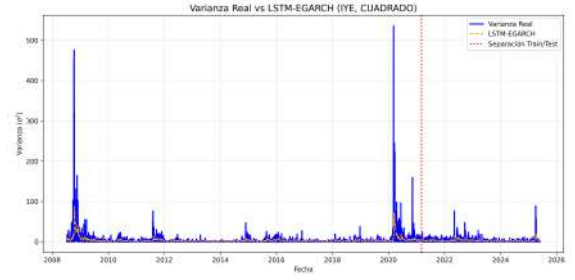
(c) Varianza - GJR-GARCH



(d) Varianza - LSTM GJR-GARCH



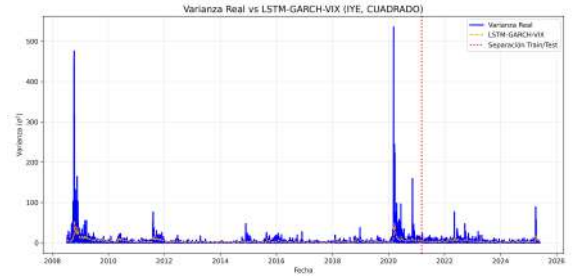
(e) Varianza - EGARCH



(f) Varianza - LSTM EGARCH



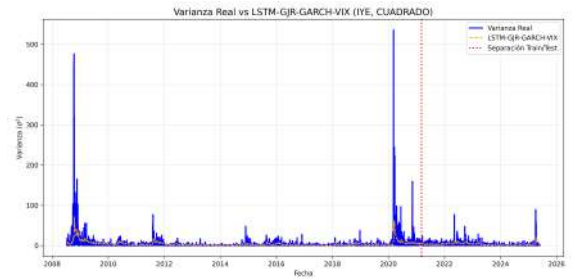
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 7: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF IYE (σ^2).

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	5,345	36,658	3,152	783949,984	127,356	1,974	0,283
GJR-GARCH	4,825	36,859	3,122	805418,894	120,312	1,973	0,413
EGARCH	6,631	36,160	3,076	598051,246	115,260	1,963	0,425
GARCH VIX	6,241	36,311	3,115	689002,417	122,727	1,952	0,308
GJR-GARCH VIX	4,282	37,069	3,124	761199,173	118,652	1,972	0,415
LSTM GARCH	7,411	35,858	2,957	437607,439	112,610	1,968	0,437
LSTM GJR-GARCH	7,348	35,882	3,068	542509,111	124,811	1,990	0,500
LSTM EGARCH	7,965	35,643	2,938	437959,113	103,924	1,998	0,498
LSTM GARCH VIX	10,117	34,809	3,153	698887,347	131,092	1,918	0,449
LSTM GJR-GARCH VIX	10,456	34,678	3,059	691159,481	119,433	1,943	0,499

Cuadro 29: Resultados de métricas de desempeño IYE (σ^2).

Modelo	Posición	Estadístico	valor-p
LSTM GARCH VIX	1	-1,8226	1
LSTM GJR-GARCH VIX	2	-1,73009	1
LSTM EGARCH	3	-0,53291	1
LSTM GARCH	4	-0,22917	1
LSTM GJR-GARCH	5	-0,21402	1
EGARCH	6	0,364421	0,9936
GARCH VIX	7	0,969734	0,7182
GJR-GARCH	8	1,168407	0,5602
GJR-GARCH VIX	9	1,394747	0,3994
GARCH	10	1,651565	0,2674

Cuadro 30: Conjunto superior de modelos para varianza de IYE.

5.4.4. Entrenamiento modelos familia GARCH IYE (σ)

El proceso de entrenamiento de la familia GARCH aplicada a la desviación estándar de la serie IYE muestra un desempeño general estable, aunque con una capacidad explicativa moderada. Los valores de R^2 % se mantienen entre 18 y 24 %, lo que indica que, si bien los modelos capturan parte de la dinámica de la volatilidad, aún persiste un componente no explicado que podría estar asociado a variaciones no lineales o factores externos.

La Tabla 31 evidencia que el modelo EGARCH alcanza los mejores resultados dentro de la familia, presentando el menor MSE y MAE, además del valor más alto de R^2 %. Asimismo, presenta un equilibrio en las métricas HMSE, HMAE y QLIKE, lo que sugiere que logra una estimación más eficiente de la varianza condicional, capturando las asimetrías típicas de la volatilidad financiera. Por su parte, los modelos que incluyen la variable VIX no presentan mejoras notables respecto a los tradicionales, lo que sugiere que el efecto de la volatilidad implícita del mercado no es determinante para explicar la dinámica de la desviación estándar en el activo IYE durante el periodo analizado.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	18,141	2,049	0,962	82,592	3,388	1,167	0,278
GJR-GARCH	20,452	1,991	0,948	80,748	3,334	1,161	0,404
EGARCH	24,227	1,897	0,933	80,054	3,323	1,160	0,424
GARCH VIX	21,059	1,976	0,949	80,835	3,369	1,165	0,319
GJR-GARCH VIX	19,524	2,014	0,949	81,773	3,335	1,160	0,393

Cuadro 31: Métricas de desempeño en entrenamiento familia GARCH IYE (σ).

5.4.5. Entrenamiento modelos híbridos Desviación Estándar IYE

Los modelos híbridos LSTM-GARCH para la desviación estándar de IYE presentan estructuras diversas y, en general, más profundas que las observadas en otras series. En la Tabla 32 se aprecia que los modelos tienden a incorporar un alto número de neuronas por capa (hasta 60), con valores de descarte aleatorio nulos o muy bajos, lo que sugiere que las redes se beneficiaron de una mayor capacidad de aprendizaje secuencial sin incurrir en sobreajuste. Además, todos los modelos utilizan una tasa de aprendizaje uniforme de 0,01, lo que permite una convergencia rápida y estable durante el entrenamiento.

Modelo LSTM	Rezagos	Neuro- nas C1	Neuro- nas C2	Neuro- nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	4	20	20	20	0,4	0,01	64	30
GJR-GARCH	15	20	8	24	0	0,01	64	70
EGARCH	4	8	2	30	0	0,01	64	60
GARCH VIX	2	30	50	50	0	0,01	64	50
GJR-GARCH VIX	13	2	60	30	0	0,01	128	70

Cuadro 32: Hiperparámetros óptimos para desviación estándar de IYE.

Los resultados del entrenamiento muestran un aumento sustancial en la capacidad explicativa en comparación con los modelos clásicos. Los valores de $R^2\%$ superan el 27 %, con reducciones consistentes en MSE y MAE. El modelo LSTM EGARCH logra el mejor ajuste en la mayoría de las métricas, mientras que los modelos que incluyen la variable VIX destacan por su eficiencia en las métricas ajustadas por heterocedasticidad (HMSE y HMAE) y en la predicción direccional (MDA). En particular, el modelo LSTM GARCH VIX alcanza el menor MSE del conjunto, acompañado de un desempeño estable en QLIKE, lo que sugiere una buena capacidad de generalización.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	27,207	1,822	0,835	49,038	2,636	1,141	0,414
LSTM GJR-GARCH	26,325	1,844	0,850	53,684	2,772	1,145	0,365
LSTM EGARCH	30,410	1,742	0,832	50,063	2,653	1,134	0,347
LSTM GARCH VIX	31,939	1,704	0,809	43,806	2,401	1,129	0,468
LSTM GJR-GARCH VIX	28,439	1,791	0,818	42,651	2,454	1,133	0,479

Cuadro 33: Métricas de desempeño en entrenamiento modelos híbridos IYE (σ).

5.4.6. Comparación de modelos fuera de muestra para Desviación Estándar IYE

Los resultados fuera de muestra, presentados en la Tabla 34, muestran que los modelos híbridos mantienen un desempeño estable y superan consistentemente a los clásicos en la mayoría de las métricas. El modelo LSTM GJR-GARCH VIX se posiciona como el de mejor rendimiento global, alcanzando los menores valores de error en todas las métricas y la mayor capacidad direccional (MDA). Esta estabilidad frente a la disminución de la bondad de ajuste observada tras el entrenamiento indica una buena capacidad de generalización, lo que sugiere que los modelos híbridos logran adaptarse a los cambios estructurales de la volatilidad de manera más eficaz que los modelos

puramente econométricos.

Si bien el modelo EGARCH clásico conserva un desempeño competitivo, destacando en la métrica MDA frente a la mayoría de los modelos, los resultados confirman que las arquitecturas híbridas superan su rendimiento en la estimación de la desviación estándar fuera de muestra, especialmente en MSE, MAE y QLIKE. Esto demuestra que la incorporación de memoria recurrente a la estructura GARCH permite capturar dependencias temporales más prolongadas, mejorando la robustez del modelo ante la volatilidad dinámica de IYE.

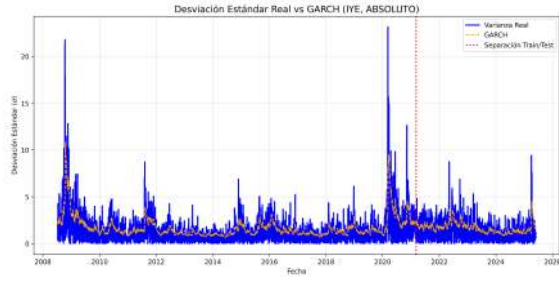
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-8,828	1,429	0,957	120,031	3,701	1,244	0,283
GJR-GARCH	-7,239	1,408	0,942	113,238	3,578	1,238	0,413
EGARCH	-5,423	1,384	0,938	108,181	3,579	1,238	0,425
GARCH VIX	-6,754	1,402	0,949	115,491	3,657	1,240	0,308
GJR-GARCH VIX	-7,552	1,412	0,942	111,601	3,567	1,238	0,415
LSTM GARCH	7,684	1,212	0,819	65,804	2,752	1,213	0,413
LSTM GJR-GARCH	7,457	1,215	0,830	71,192	2,882	1,214	0,393
LSTM EGARCH	8,674	1,199	0,820	66,040	2,773	1,209	0,351
LSTM GARCH VIX	10,091	1,180	0,803	58,369	2,565	1,205	0,462
LSTM GJR-GARCH VIX	11,625	1,160	0,796	54,183	2,524	1,203	0,477

Cuadro 34: Resultados de métricas de desempeño IYE (σ).

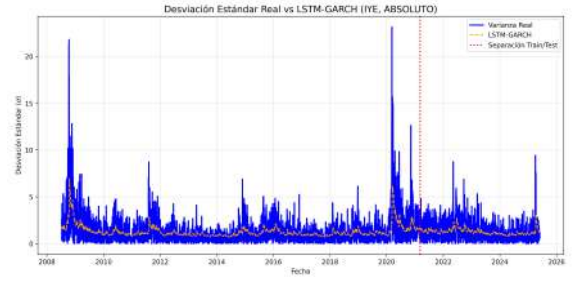
La prueba de conjunto superior respalda la superioridad estadística de los modelos híbridos frente a los clásicos. Dicha prueba elimina todas las variantes tradicionales y el modelo LSTM GJR-GARCH, confirmando que las mejoras observadas en las métricas de la Tabla 34 son significativas al 5 % de confianza.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,76448	1
LSTM GARCH VIX	2	-1,02662	1
LSTM GARCH	3	1,959608	0,0892
LSTM EGARCH	4	2,038844	0,074

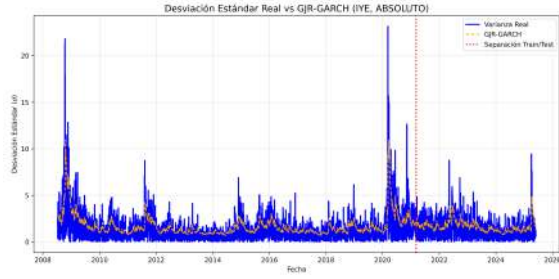
Cuadro 35: Conjunto superior de modelos para desviación estándar de IYE.



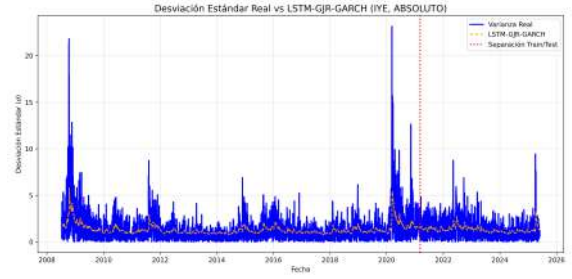
(a) Desviación estándar - GARCH



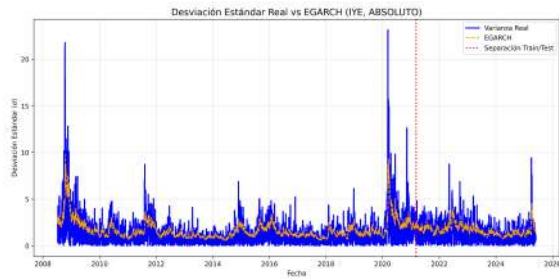
(b) Desviación estándar - LSTM GARCH



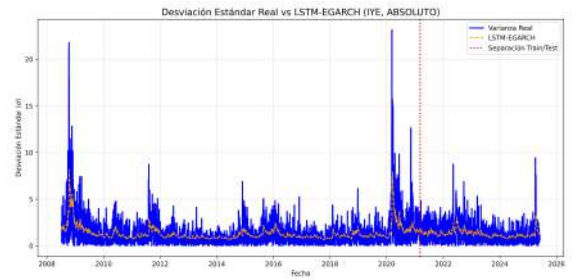
(c) Desviación estándar - GJR-GARCH



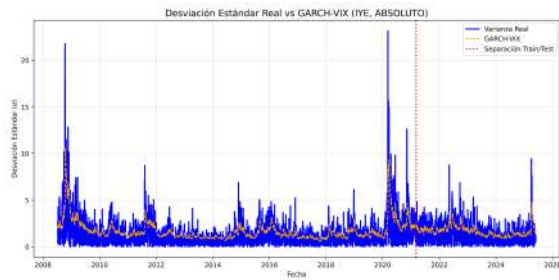
(d) Desviación estándar - LSTM GJR-GARCH



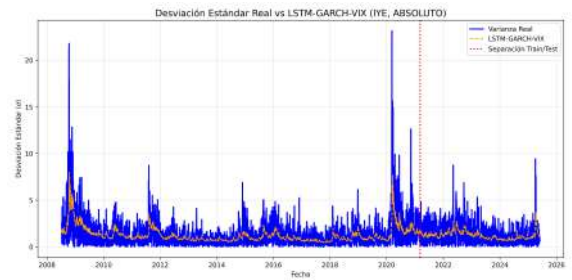
(e) Desviación estándar - EGARCH



(f) Desviación estándar - LSTM EGARCH



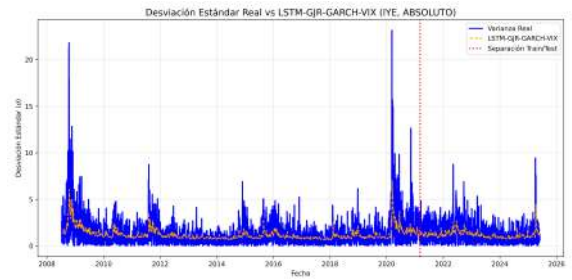
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.5. PBW

5.5.1. Entrenamiento modelos familia GARCH PBW (σ^2)

La optimización mediante la matriz de rezagos arrojó los resultados que se muestran en la Tabla 36. En general, los modelos GARCH sin variable exógena presentaron especificaciones de baja complejidad, predominando configuraciones con valores de p y q en torno a dos. El modelo con mejor desempeño a nivel de AIC corresponde al GJR-GARCH con la variable VIX, lo que sugiere que la incorporación de la volatilidad implícita del mercado contribuye a mejorar la capacidad explicativa del modelo sobre la varianza condicional. Asimismo, los modelos GJR-GARCH simples presentan un ajuste competitivo, confirmando la presencia de asimetrías en la respuesta de la volatilidad frente a shocks negativos.

Modelo	p	o	q	vix	AIC
GARCH	2	-	2	-	12882,949
GJR-GARCH	2	1	2	-	12866,635
EGARCH	1	2	1	-	12889,956
GARCH VIX	2	2	2	2	12899,442
GJR-GARCH VIX	5	5	1	5	12852,387

Cuadro 36: Especificación de rezagos óptimos para modelos de familia GARCH PBW.

En términos de desempeño, los resultados de la Tabla 37 indican que el modelo EGARCH logra el mejor equilibrio global, alcanzando el menor MSE y MAE junto con una capacidad direccional (MDA) de 0,473, la más alta entre los modelos clásicos. Este resultado sugiere que la estructura exponencial logra capturar mejor las asimetrías en la volatilidad del activo. Los modelos GJR-GARCH y GJR-GARCH VIX también muestran un rendimiento competitivo, lo que refuerza la presencia del efecto apalancamiento en la serie PBW. En contraste, los modelos simples GARCH y GARCH VIX exhiben un desempeño más débil, con errores heterocedásticos significativamente mayores, lo que indica limitaciones en su capacidad de adaptación frente a cambios bruscos en la varianza condicional.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	24,003	174,928	5,147	3514898,779	74,318	2,239	0,417
GJR-GARCH	25,357	171,812	5,103	1245556,829	57,213	2,234	0,449
EGARCH	23,771	175,464	5,057	796843,593	51,124	2,230	0,473
GARCH VIX	24,038	174,849	5,148	3525023,288	74,256	2,239	0,419
GJR-GARCH VIX	25,429	171,647	5,115	1421067,901	57,905	2,223	0,434

Cuadro 37: Métricas de desempeño en entrenamiento familia GARCH PBW (σ^2).

5.5.2. Entrenamiento modelos híbridos Varianza PBW

Los modelos híbridos LSTM-GARCH presentan un comportamiento heterogéneo, tanto en arquitectura como en desempeño. De acuerdo con la Tabla 38, las redes neuronales utilizan entre dos y tres rezagos, con un número de neuronas intermedio a alto y tasas de aprendizaje entre 0,001 y 0,01. Se observa que los modelos que integran la variable VIX tienden a tener mayor profundidad (mayor cantidad de neuronas) y un descarte aleatorio reducido o nulo, lo que les permite capturar dependencias temporales más prolongadas en la volatilidad.

Modelo LSTM	Rezagos	Neuro- nas C1	Neuro- nas C2	Neuro- nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	3	2	4	16	0,4	0,01	64	50
GJR-GARCH	3	24	16	20	0,4	0,001	128	70
EGARCH	3	24	16	20	0,4	0,001	128	70
GARCH VIX	3	50	24	16	0	0,01	64	60
GJR-GARCH VIX	2	4	40	2	0,2	0,01	64	60

Cuadro 38: Hiperparámetros óptimos para varianza de PBW.

Los resultados de entrenamiento muestran que los modelos híbridos no superan de manera sistemática a sus contrapartes clásicas, aunque sí logran un rendimiento competitivo en la mayoría de las métricas. En particular, el modelo LSTM GARCH VIX alcanza el mayor $R^2\%$ y el menor MSE, lo que indica una mejor capacidad de ajuste durante el entrenamiento. Sin embargo, el aumento en las métricas HMSE y HMAE en algunos modelos híbridos sugiere la presencia de cierta sensibilidad ante observaciones extremas. El modelo LSTM EGARCH, en cambio, mantiene un equilibrio más estable entre precisión y robustez, con un QLIKE menor que la mayoría de las configuraciones evaluadas.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	25,299	171,946	4,852	1399720,390	54,948	2,240	0,442
LSTM GJR-GARCH	24,789	173,119	5,302	970843,267	59,276	2,266	0,463
LSTM EGARCH	24,317	174,207	4,973	823349,760	51,047	2,245	0,454
LSTM GARCH VIX	27,137	167,716	5,818	1907764,152	76,939	2,297	0,466
LSTM GJR-GARCH VIX	22,325	178,792	6,947	2148230,530	98,076	2,484	0,475

Cuadro 39: Métricas de desempeño en entrenamiento modelos híbridos PBW (σ^2).

5.5.3. Comparación de modelos fuera de muestra para Varianza PBW

La Tabla 40 presenta los resultados de desempeño fuera de muestra, donde se observa una tendencia general a la mejora del ajuste por parte de los modelos híbridos, especialmente en las métricas de error cuadrático medio y capacidad direccional. El modelo LSTM GARCH VIX destaca como el más eficiente en términos globales, alcanzando el menor MSE y los mejores valores en QLIKE y MDA. Esto sugiere que la combinación de la estructura LSTM con la información del índice VIX permite capturar tanto los patrones secuenciales de la volatilidad como los choques exógenos del mercado.

Por su parte, los modelos EGARCH y GJR-GARCH VIX conservan un rendimiento competitivo, especialmente en la predicción direccional, lo que refleja su capacidad para modelar el efecto apalancamiento y la asimetría de los retornos. Sin embargo, los resultados de HMSE y HMAE muestran que algunos modelos, en particular los híbridos más complejos, son sensibles a la presencia de valores extremos, lo cual podría relacionarse con la elevada varianza del ETF PBW.

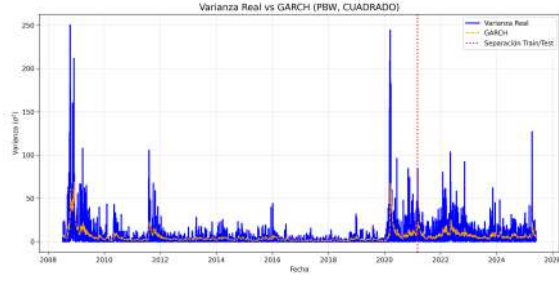
Modelo	R^2 %	MSE %²	MAE %²	HMSE %²	HMAE %²	QLIKE	MDA
GARCH	3,661	129,737	7,215	1.402.980,810	126,261	3,005	0,386
GJR-GARCH	4,623	128,442	7,152	1.473.281,215	125,791	3,006	0,441
EGARCH	6,131	126,411	7,420	1.611.554,781	133,356	3,000	0,476
GARCH-VIX	3,654	129,746	7,221	1.405.840,617	126,465	3,005	0,389
GJR-GARCH-VIX	7,171	125,010	7,141	1.275.159,005	121,162	3,000	0,429
LSTM-GARCH	5,263	127,580	6,895	962.207,830	102,926	3,044	0,453
LSTM-GJR-GARCH	6,579	125,808	7,157	1.797.046,727	128,309	2,991	0,504
LSTM-EGARCH	5,805	126,850	7,014	1.556.555,692	120,786	3,003	0,493
LSTM-GARCH-VIX	8,825	122,783	6,981	1.168.093,978	114,381	2,997	0,468
LSTM-GJR-GARCH-VIX	7,855	124,089	7,479	2.101.100,038	148,935	2,979	0,468

Cuadro 40: Resultados de métricas de desempeño para la varianza de PBW (σ^2).

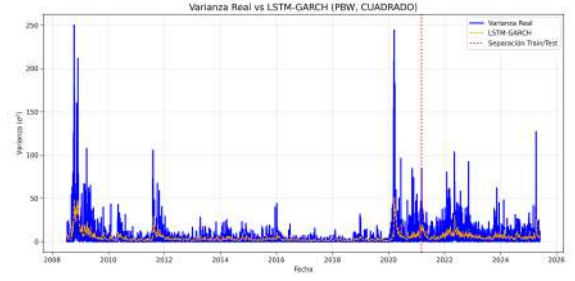
Finalmente, aunque los modelos híbridos exhiben un desempeño superior en la mayoría de las métricas, la prueba del conjunto superior de modelos (MCS), presentada en la Tabla 41, indica que dichas mejoras no son estadísticamente significativas. Esto se debe a que ninguno de los diez modelos comparados fue eliminado del conjunto, lo que sugiere que sus diferencias en precisión predictiva no son concluyentes desde un punto de vista estadístico.

Modelo	Posición	Estadístico	valor-p
LSTM-GARCH-VIX	1	-2,139	1,000
GJR-GARCH-VIX	2	-1,417	1,000
LSTM-GJR-GARCH-VIX	3	-1,372	1,000
LSTM-GJR-GARCH	4	-1,114	1,000
EGARCH	5	-0,161	1,000
LSTM-EGARCH	6	0,190	1,000
LSTM-GARCH	7	1,249	0,633
GJR-GARCH	8	1,429	0,503
GARCH	9	1,941	0,185
GARCH-VIX	10	1,952	0,182

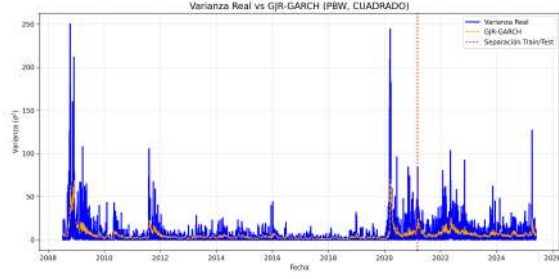
Cuadro 41: Conjunto superior de modelos para varianza PBW.



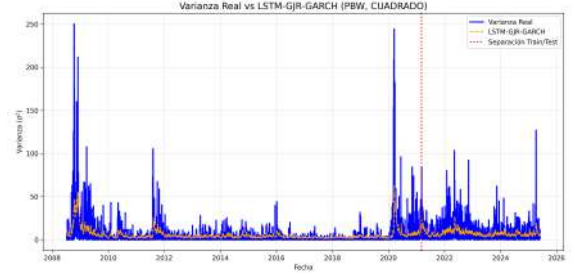
(a) Varianza - GARCH



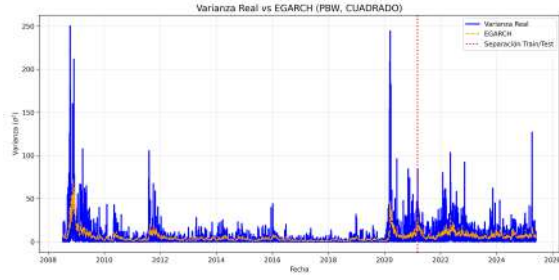
(b) Varianza - LSTM GARCH



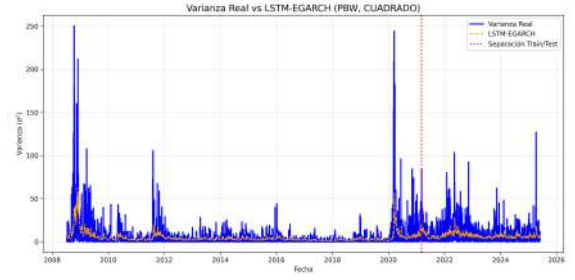
(c) Varianza - GJR-GARCH



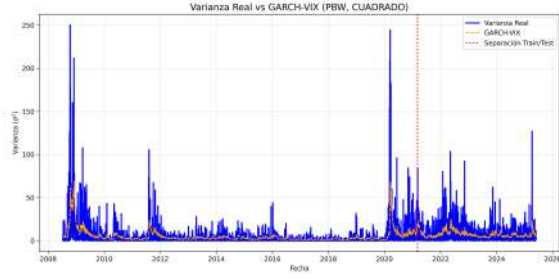
(d) Varianza - LSTM GJR-GARCH



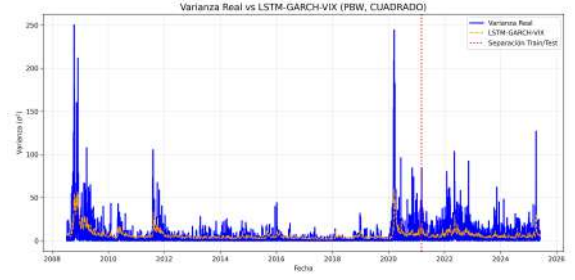
(e) Varianza - EGARCH



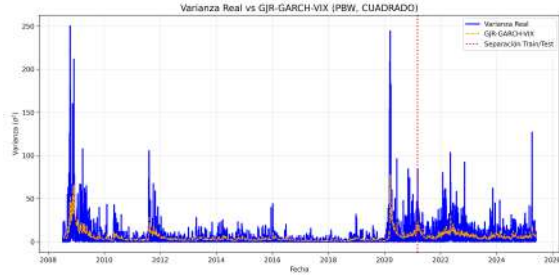
(f) Varianza - LSTM EGARCH



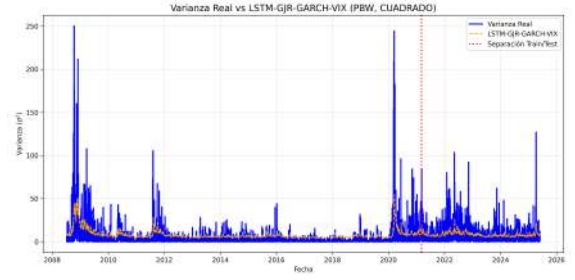
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 9: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF PBW (σ^2).

5.5.4. Entrenamiento modelos familia GARCH PBW (σ)

Los resultados aplicar raíz cuadrada a los modelos de la familia GARCH ajustados a la varianza del ETF PBW se presentan en la Tabla 42. En general, los modelos presentan un desempeño moderado, con coeficientes de determinación cercanos al 20% y valores de MSE en torno a 2,1. El modelo EGARCH logra los mejores resultados globales, obteniendo el menor MSE y MAE junto con la mayor capacidad direccional ($MDA = 0,473$), lo que indica que su estructura logarítmica es más eficaz para capturar las asimetrías y la dinámica no lineal de la volatilidad del activo.

Por su parte, los modelos GJR-GARCH y GJR-GARCH VIX también muestran un comportamiento sólido, superando en precisión a la versión estándar del GARCH. La inclusión de la variable VIX mejora ligeramente la estabilidad del modelo ante shocks de volatilidad, aunque su efecto es limitado. En contraste, los modelos simples GARCH y GARCH VIX presentan un rendimiento casi idéntico, evidenciando que el componente exógeno no aporta una mejora sustancial en el contexto de la desviación estándar. En conjunto, se observa que la estructura EGARCH ofrece la mejor relación entre ajuste y parsimonia dentro de la familia clásica.

Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	19,825	2,183	1,082	69,791	2,301	0,372	0,417
GJR-GARCH	21,167	2,146	1,074	52,785	2,252	0,370	0,449
EGARCH	21,893	2,126	1,069	46,739	2,230	0,369	0,473
GARCH VIX	19,838	2,182	1,082	69,734	2,299	0,372	0,419
GJR-GARCH VIX	21,571	2,135	1,072	53,503	2,239	0,367	0,434

Cuadro 42: Métricas de desempeño en entrenamiento familia GARCH PBW (σ).

5.5.5. Entrenamiento modelos híbridos Desviación Estándar PBW

La Tabla 43 presenta los hiperparámetros óptimos empleados en los modelos híbridos LSTM-GARCH. Se observa que las arquitecturas seleccionadas son relativamente profundas, utilizando entre dos y seis rezagos y capas densas con hasta cincuenta neuronas, lo que refleja una alta capacidad de representación temporal. Las tasas de aprendizaje son consistentes en 0,01, mientras que el descarte aleatorio (dropout) varía entre 0 y 0,4, lo que indica un nivel de regularización moderado destinado a evitar sobreajuste.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	3	24	12	8	0	0,01	64	30
GJR-GARCH	3	12	8	24	0,2	0,01	64	60
EGARCH	6	30	30	30	0	0,01	128	40
GARCH VIX	4	16	50	30	0,2	0,01	64	70
GJR-GARCH VIX	2	20	30	24	0,4	0,01	128	40

Cuadro 43: Hiperparámetros óptimos para desviación estándar de PBW.

Los resultados de la Tabla 44 evidencian una mejora general de los modelos híbridos respecto a sus contrapartes clásicas. Los valores de $R^2\%$ aumentan hasta un 28,7% en el modelo LSTM GARCH VIX, que además alcanza el menor MSE y MAE del conjunto. Este resultado sugiere que la combinación entre las dependencias no lineales capturadas por la red LSTM y la estructura autoregresiva del GARCH permite modelar de forma más efectiva los patrones de volatilidad.

El modelo LSTM EGARCH mantiene una relación equilibrada entre precisión y robustez, con las métricas HMSE y HMAE más bajas del grupo, indicando una menor sensibilidad a valores extremos. Por su parte, el modelo LSTM GJR-GARCH VIX, aunque no alcanza los mejores valores de error, conserva una capacidad direccional competitiva, lo que sugiere que el efecto apalancamiento sigue siendo relevante para explicar la dinámica de la volatilidad diaria del ETF PBW.

Modelo	$R^2\%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	24,191	2,064	1,031	43,971	2,116	0,368	0,437
LSTM GJR-GARCH	26,332	2,006	0,985	35,365	1,906	0,354	0,412
LSTM EGARCH	28,134	1,956	0,986	33,797	1,923	0,354	0,362
LSTM GARCH VIX	28,689	1,941	0,982	34,383	1,890	0,347	0,446
LSTM GJR-GARCH VIX	27,040	1,986	1,013	43,876	2,056	0,355	0,433

Cuadro 44: Métricas de desempeño en entrenamiento modelos híbridos PBW (σ).

5.5.6. Comparación de modelos fuera de muestra para Desviación Estándar PBW

La evaluación fuera de muestra, presentada en la Tabla 45, muestra una clara estabilidad en el comportamiento de los modelos híbridos, los cuales logran mantener un desempeño superior en la mayoría de las métricas. En particular, los modelos LSTM-GARCH y LSTM-GJR-GARCH alcanzan mejoras en el MSE y MAE, junto con incrementos notables en la capacidad direccional

(MDA), superando consistentemente a los modelos clásicos.

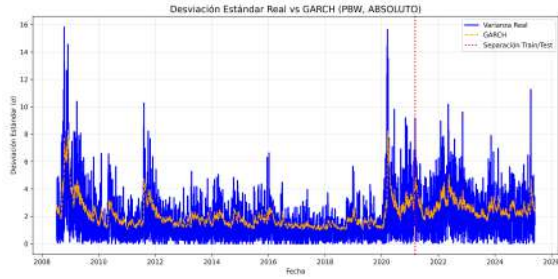
El modelo LSTM GJR-GARCH VIX se destaca por obtener los mejores resultados globales, con la mayor bondad de ajuste ($R^2 \% = 5,08$) y los menores errores heterocedásticos, lo que demuestra su capacidad para capturar de forma precisa la dinámica no lineal y asimétrica de la volatilidad. Los modelos con la variable VIX tienden a presentar un mejor equilibrio entre ajuste y robustez, mientras que los modelos EGARCH, tanto clásicos como híbridos, continúan mostrando un comportamiento estable pero con menor poder predictivo fuera del entrenamiento.

En términos generales, los resultados reflejan que los modelos híbridos ofrecen una mejor generalización ante condiciones de mercado cambiantes, manteniendo bajos niveles de error y un rendimiento direccional competitivo.

Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-6,821	2,949	1,405	120,267	3,039	1,799	0,386
GJR-GARCH	-5,219	2,905	1,394	119,868	3,005	1,797	0,441
EGARCH	-9,550	3,024	1,445	127,133	3,149	1,806	0,476
GARCH-VIX	-6,931	2,952	1,406	120,465	3,042	1,799	0,389
GJR-GARCH-VIX	-4,326	2,880	1,397	115,257	2,995	1,797	0,429
LSTM-GARCH	3,552	2,662	1,288	79,704	2,472	1,782	0,469
LSTM-GJR-GARCH	4,012	2,650	1,271	74,493	2,354	1,782	0,447
LSTM-EGARCH	3,774	2,656	1,286	80,240	2,430	1,783	0,356
LSTM-GARCH-VIX	3,236	2,671	1,264	64,096	2,160	1,791	0,412
LSTM-GJR-GARCH-VIX	5,078	2,620	1,283	77,546	2,392	1,782	0,439

Cuadro 45: Resultados de métricas de desempeño para el ETF PBW (σ).

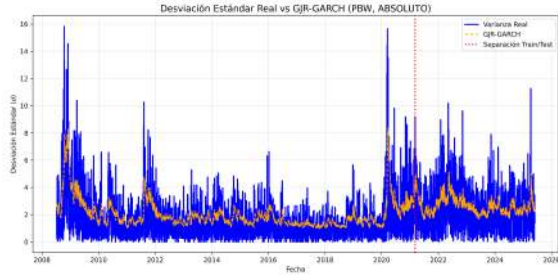
Los modelos híbridos establecen una superioridad estadísticamente significativa frente a los modelos clásicos, según los resultados de la prueba del conjunto superior de modelos (MCS) presentada en la Tabla 46. No obstante, se observa que los modelos híbridos que incorporan la variable VIX no superan consistentemente a sus contrapartes sin dicha variable en términos de error cuadrático medio (MSE), lo que sugiere que la inclusión de la volatilidad implícita no aporta mejoras sustanciales en la predicción de la desviación estándar dentro de este contexto.



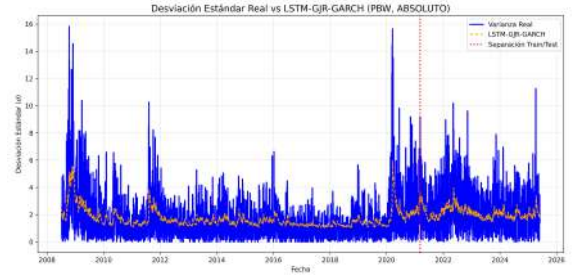
(a) Desviación estándar - GARCH



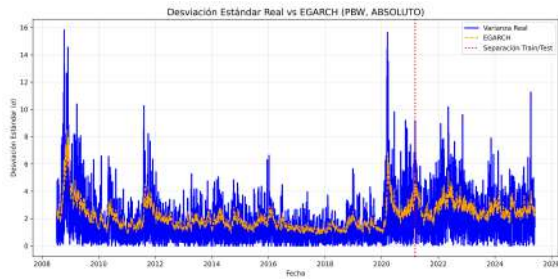
(b) Desviación estándar - LSTM GARCH



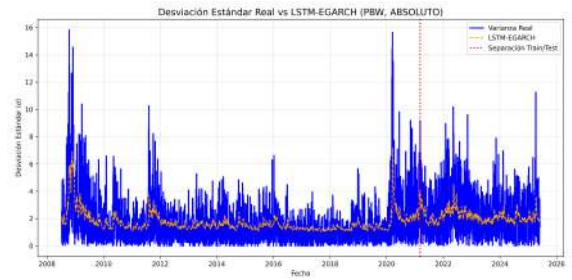
(c) Desviación estándar - GJR-GARCH



(d) Desviación estándar - LSTM GJR-GARCH



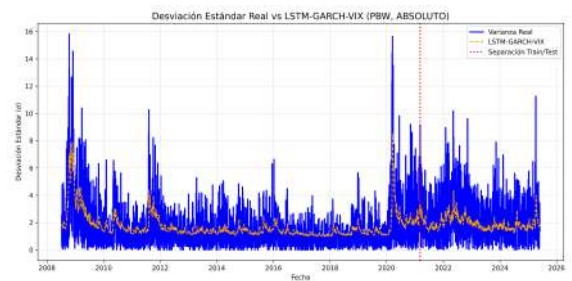
(e) Desviación estándar - EGARCH



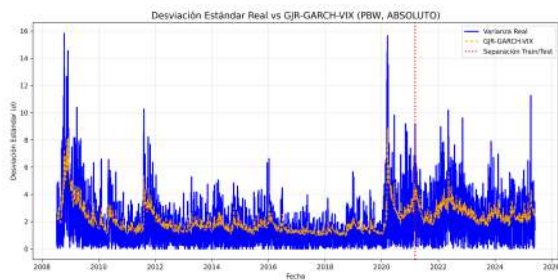
(f) Desviación estándar - LSTM EGARCH



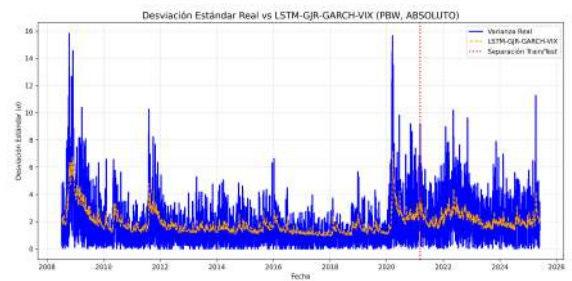
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

Modelo	Posición	Estadístico	valor-p
LSTM-GJR-GARCH-VIX	1	-1,84664	1
LSTM-GJR-GARCH	2	-0,19287	1
LSTM-EGARCH	3	0,393228	0,9614
LSTM-GARCH	4	0,706849	0,7988
LSTM-GARCH-VIX	5	1,064402	0,5474

Cuadro 46: Conjunto superior de modelos para desviación estándar PBW.

5.6. QCLN

5.6.1. Entrenamiento modelos familia GARCH QCLN (σ^2)

La optimización de los rezagos, presentada en la Tabla 47, evidencia que los modelos sin variable exógena tienden a presentar estructuras más compactas, mientras que aquellos que incorporan la variable VIX requieren configuraciones más complejas. En general, el modelo GJR-GARCH de orden (3,3,1) se posiciona como el mejor dentro de la familia tradicional, alcanzando el menor valor de AIC. Este resultado sugiere la existencia de asimetrías significativas en el comportamiento de la volatilidad, donde los retornos negativos tienden a generar incrementos más pronunciados en la varianza condicional.

Modelo	p	o	q	vix	AIC
GARCH	1	-	1	-	12604,413
GJR-GARCH	3	3	1	-	12577,381
EGARCH	2	4	5	-	12597,116
GARCH VIX	4	-	3	5	12598,280
GJR-GARCH VIX	3	3	5	1	12592,369

Cuadro 47: Especificación de rezagos óptimos para modelos de familia GARCH QCLN.

En términos de desempeño, los resultados de la Tabla 48 muestran que los modelos asimétricos obtienen mejores métricas que las versiones simétricas. El modelo GJR-GARCH VIX alcanza el mayor coeficiente de determinación ($R^2 = 24,76\%$), el menor MSE y los valores más bajos de HMSE y HMAE, lo que indica un ajuste más preciso de la varianza condicional. Por su parte, el modelo EGARCH destaca por su mejor desempeño direccional ($MDA = 0,488$), reflejando una mayor capacidad para anticipar los movimientos de volatilidad.

Estos resultados demuestran que las estructuras que capturan no linealidades y efectos de apalancamiento —como GJR-GARCH y EGARCH— son más adecuadas para representar la dinámica del ETF QCLN, especialmente frente a fluctuaciones abruptas del mercado o períodos de alta volatilidad.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	21,946	144,516	4,658	123334,218	53,551	2,129	0,287
GJR-GARCH	24,002	140,710	4,606	119724,872	52,859	2,121	0,399
EGARCH	22,739	143,049	4,577	100970,584	51,579	2,115	0,488
GARCH VIX	23,261	142,083	4,603	96358,452	51,231	2,121	0,388
GJR-GARCH VIX	24,761	139,305	4,513	84470,459	49,220	2,104	0,432

Cuadro 48: Métricas de desempeño en entrenamiento familia GARCH QCLN (σ^2).

5.6.2. Entrenamiento modelos híbridos Varianza QCLN

Los hiperparámetros óptimos utilizados en los modelos híbridos LSTM-GARCH se presentan en la Tabla 49. Las configuraciones seleccionadas muestran una preferencia por arquitecturas moderadamente profundas, con tres rezagos predominantes y entre 4 y 60 neuronas por capa, dependiendo del modelo. Las tasas de aprendizaje utilizadas (0,001 y 0,01) indican un proceso de ajuste controlado, orientado a maximizar la estabilidad y evitar oscilaciones durante la optimización. Los modelos que incluyen la variable VIX tienden a emplear menos rezagos, pero un mayor número de neuronas, compensando la menor dependencia temporal con un aprendizaje más complejo sobre los efectos exógenos del mercado.

Modelo LSTM	Rezagos	Neuronas C1	Neuronas C2	Neuronas C3	Descarte Aleatorio	Tasa Aprendizaje	Lote	Épocas
GARCH	3	24	2	30	0,2	0,01	64	30
GJR-GARCH	3	24	24	16	0,4	0,01	128	70
EGARCH	3	24	4	16	0,4	0,01	128	50
GARCH VIX	1	60	2	60	0,2	0,001	64	40
GJR-GARCH VIX	2	4	4	12	0,2	0,01	64	60

Cuadro 49: Hiperparámetros óptimos para varianza de QCLN.

Los resultados de entrenamiento (Tabla 50) muestran una mejora leve pero consistente en el desempeño de los modelos híbridos respecto a los modelos tradicionales. El modelo LSTM GARCH VIX presenta el mayor R^2 (25,37 %) y el menor MSE, consolidándose como el más eficiente en la fase de ajuste. Los modelos LSTM EGARCH y LSTM GJR-GARCH exhiben un desempeño competitivo, aunque con un incremento en HMAE y QLIKE, lo que sugiere una ligera pérdida de eficiencia en presencia de observaciones extremas.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	22,116	144,202	4,775	107740,364	59,826	2,154	0,462
LSTM GJR-GARCH	21,403	145,522	4,992	95513,203	65,473	2,218	0,461
LSTM EGARCH	22,557	143,386	4,899	84251,645	61,767	2,196	0,460
LSTM GARCH VIX	25,370	138,177	4,985	159018,958	60,382	2,189	0,492
LSTM GJR-GARCH VIX	24,311	140,137	5,740	178219,315	83,740	2,275	0,485

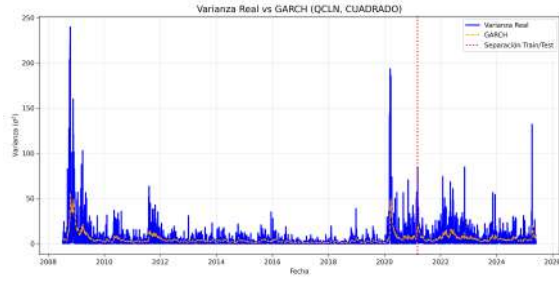
Cuadro 50: Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ^2).

5.6.3. Comparación de modelos fuera de muestra para Varianza QCLN

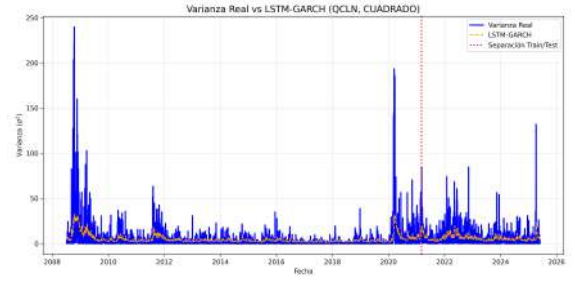
La comparación fuera de muestra, presentada en la Tabla 51, confirma que los modelos híbridos mantienen un rendimiento superior al de las versiones clásicas. El modelo LSTM GARCH VIX obtiene los mejores resultados globales, con un R^2 de 8,83 % y el menor MSE (92,573), lo que evidencia una mejor capacidad de generalización frente a los cambios de régimen en la volatilidad. De manera similar, los modelos LSTM EGARCH y LSTM GJR-GARCH muestran un equilibrio sólido entre ajuste y estabilidad, alcanzando las menores métricas de error absoluto y valores altos de MDA.

Los modelos clásicos, en particular EGARCH y GJR-GARCH, conservan un desempeño aceptable, aunque se ven superados en precisión y robustez por sus equivalentes híbridos. Las métricas HMSE y HMAE indican que las versiones LSTM reducen la magnitud de los errores extremos, lo que sugiere una mayor capacidad de adaptación a eventos de alta volatilidad. Estos resultados reflejan la ventaja de incorporar componentes de memoria larga en la modelación de la varianza condicional.

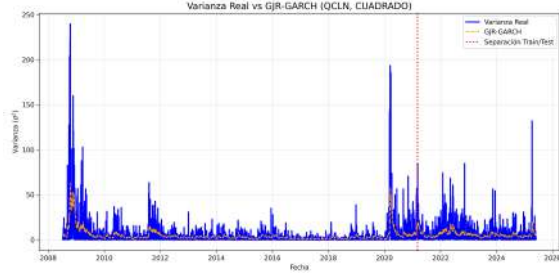
La prueba del conjunto superior de modelos (MCS), presentada en la Tabla 52, indica que la aparente superioridad de los modelos híbridos frente a la mayoría de los modelos clásicos resulta ser marginal. En efecto, los resultados sugieren que las arquitecturas híbridas no logran dominar de forma estadísticamente significativa —al menos en términos del error cuadrático medio (MSE)— a sus contrapartes econométricas tradicionales. Esto evidencia que, si bien los modelos híbridos pueden ofrecer un mejor ajuste numérico en ciertas métricas, dichas mejoras no alcanzan un nivel de significancia suficiente para ser consideradas robustas en un marco comparativo formal.



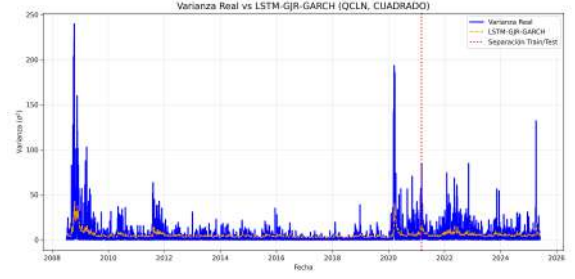
(a) Varianza - GARCH



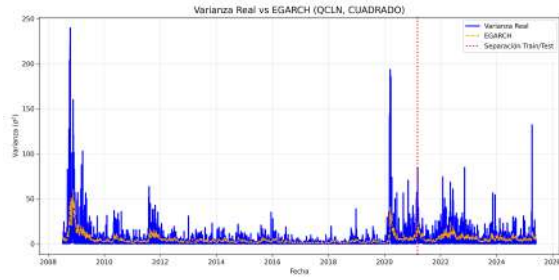
(b) Varianza - LSTM GARCH



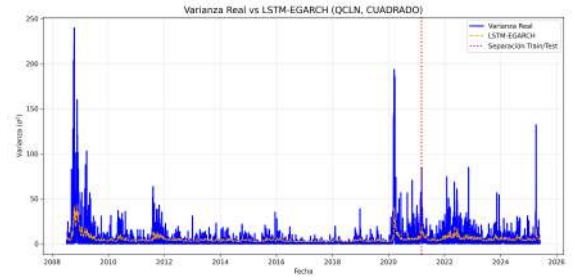
(c) Varianza - GJR-GARCH



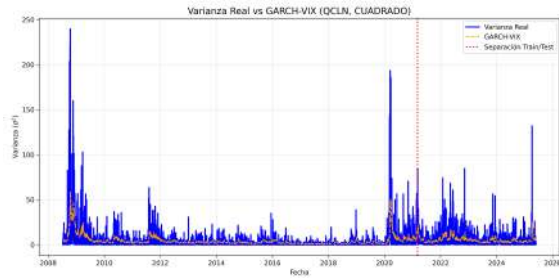
(d) Varianza - LSTM GJR-GARCH



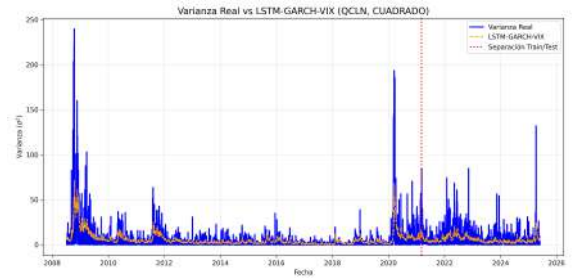
(e) Varianza - EGARCH



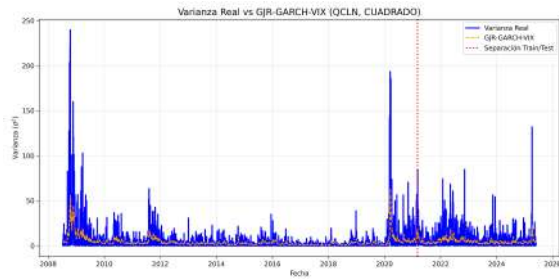
(f) Varianza - LSTM EGARCH



(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 11: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF QCLN (σ^2).

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	3,859	97,619	5,931	2133623,878	128,822	2,751	0,278
GJR-GARCH	5,972	95,474	5,820	2324725,271	128,384	2,744	0,399
EGARCH	7,474	93,949	5,989	2535884,802	140,859	2,740	0,465
GARCH VIX	3,776	97,704	5,737	1983796,592	116,516	2,772	0,369
GJR-GARCH VIX	6,499	94,939	5,509	2238563,061	114,788	2,805	0,394
LSTM GARCH	6,412	95,027	5,949	1993854,806	129,318	2,737	0,480
LSTM GJR-GARCH	6,877	94,555	5,773	1773031,184	122,301	2,731	0,475
LSTM EGARCH	7,861	93,556	5,812	2003274,960	127,754	2,728	0,471
LSTM GARCH VIX	8,829	92,573	5,655	1916473,279	115,377	2,799	0,468
LSTM GJR-GARCH VIX	9,182	92,215	6,050	2681891,479	140,361	2,729	0,462

Cuadro 51: Resultados de métricas de desempeño QCLN (σ^2).

Modelo	Posición	Estadístico	valor-p
LSTM EGARCH	1	-1,98694	1
LSTM GJR-GARCH VIX	2	-1,5375	1
LSTM GARCH VIX	3	-1,26454	1
EGARCH	4	-0,78737	1
LSTM GJR-GARCH	5	-0,41386	1
GJR-GARCH VIX	6	0,265346	0,9996
LSTM GARCH	7	0,642061	0,97
GJR-GARCH	8	0,855126	0,906
GARCH	9	2,191901	0,1074
GARCH VIX	10	2,337259	0,074

Cuadro 52: Conjunto superior de modelos para varianza de QCLN.

5.6.4. Entrenamiento modelos familia GARCH QCLN (σ)

Los resultados obtenidos para la familia GARCH, presentados en la Tabla 53, indican un desempeño relativamente homogéneo entre las especificaciones, con leves mejoras en los modelos que incorporan asimetrías o efectos apalancamiento. El modelo GJR-GARCH-VIX alcanza la mayor capacidad explicativa con un R^2 de 22,28 %, junto con los menores valores de MSE y QLIKE, lo que refleja un ajuste más eficiente de la desviación estándar condicional. Asimismo, el modelo EGARCH muestra un desempeño competitivo, obteniendo la mejor métrica MDA, lo que sugiere una mejor habilidad para anticipar la dirección de los cambios en la volatilidad.

En general, los resultados confirman que las especificaciones asimétricas capturan de mejor forma los choques negativos en los retornos, asociados a incrementos en la volatilidad. La inclusión del índice VIX mejora marginalmente las métricas de error, aunque sin producir diferencias sustanciales respecto a los modelos sin variable exógena.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	17,680	2,017	1,049	47,325	2,749	1,339	0,287
GJR-GARCH	19,492	1,972	1,037	46,628	2,726	1,336	0,399
EGARCH	20,273	1,953	1,035	45,421	2,709	1,334	0,488
GARCH VIX	19,595	1,970	1,043	45,151	2,724	1,337	0,388
GJR-GARCH VIX	22,277	1,904	1,026	43,215	2,675	1,332	0,431

Cuadro 53: Métricas de desempeño en entrenamiento familia GARCH QCLN (σ).

5.6.5. Entrenamiento modelos híbridos Desviación Estándar QCLN

Los hiperparámetros óptimos de los modelos híbridos se presentan en la Tabla 54. En general, las configuraciones muestran arquitecturas moderadas, con entre 1 y 13 rezagos, tasas de aprendizaje de 0,01 y bajos niveles de descarte aleatorio, lo que indica un ajuste enfocado en estabilidad y precisión. Los modelos que incorporan la variable VIX utilizan más neuronas y un mayor número de épocas, lo que podría estar asociado a la necesidad de capturar la volatilidad implícita proveniente de dicho índice.

En la Tabla 55, se observa una mejora generalizada de las métricas respecto a los modelos tradicionales. El LSTM GJR-GARCH destaca como el modelo más eficiente, con un R^2 de 28,21 %,

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	3	4	12	12	0,4	0,01	64	40
GJR-GARCH	13	8	20	12	0,2	0,01	64	70
EGARCH	1	8	12	4	0	0,01	128	50
GARCH VIX	3	12	50	8	0	0,01	64	70
GJR-GARCH VIX	1	12	30	20	0	0,01	128	70

Cuadro 54: Hiperparámetros óptimos para desviación estándar de QCLN.

el menor MSE y un MAE de 0,932, consolidándose como la mejor arquitectura híbrida durante la fase de entrenamiento. Asimismo, los modelos LSTM EGARCH y LSTM GARCH VIX obtienen resultados competitivos, aunque con ligeros aumentos en HMAE y QLIKE, lo que indica una leve pérdida de precisión en observaciones extremas.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	24,014	1,861	0,969	33,195	2,391	1,324	0,428
LSTM GJR-GARCH	28,214	1,759	0,932	29,974	2,202	1,310	0,438
LSTM EGARCH	25,007	1,837	0,960	32,092	2,347	1,320	0,447
LSTM GARCH VIX	25,460	1,826	0,967	32,132	2,375	1,321	0,446
LSTM GJR-GARCH VIX	25,647	1,821	0,989	37,449	2,511	1,323	0,449

Cuadro 55: Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ).

5.6.6. Comparación de modelos fuera de muestra para Desviación Estándar QCLN

La comparación fuera de muestra (Tabla 56) confirma que los modelos híbridos superan a los tradicionales en la mayoría de las métricas. El modelo LSTM GJR-GARCH VIX obtiene el mejor desempeño global, con un R^2 de 6,75 %, el menor MSE (2,172) y un MAE de 1,137. Estas mejoras se reflejan también en las métricas HMSE y HMAE, donde los modelos híbridos presentan menores valores, lo que sugiere una mayor robustez frente a eventos de alta volatilidad.

Aunque los modelos clásicos, en particular el GJR-GARCH-VIX, muestran cierto nivel de estabilidad, los resultados evidencian que las estructuras híbridas con redes LSTM logran capturar mejor los patrones persistentes de volatilidad y adaptarse a cambios en la dinámica del mercado. En términos de predicción direccional, los modelos LSTM GARCH VIX y LSTM GJR-GARCH VIX alcanzan los mayores valores de MDA, lo que indica una mejora en la capacidad de anticipar

los movimientos de la volatilidad.

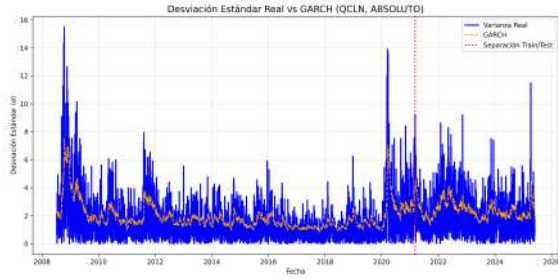
Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-6,961	2,491	1,277	123,085	2,910	1,663	0,278
GJR-GARCH	-4,274	2,429	1,258	122,721	2,873	1,659	0,399
EGARCH	-7,836	2,512	1,294	134,857	3,038	1,667	0,465
GARCH VIX	-2,677	2,392	1,239	111,151	2,731	1,656	0,369
GJR-GARCH VIX	2,256	2,277	1,198	109,687	2,606	1,650	0,394
LSTM GARCH	5,209	2,208	1,148	76,437	2,329	1,642	0,434
LSTM GJR-GARCH	5,739	2,196	1,147	83,390	2,318	1,643	0,418
LSTM EGARCH	6,033	2,189	1,148	80,969	2,371	1,641	0,429
LSTM GARCH VIX	5,261	2,207	1,124	59,530	2,042	1,645	0,460
LSTM GJR-GARCH VIX	6,753	2,172	1,137	74,567	2,209	1,643	0,430

Cuadro 56: Resultados de métricas de desempeño QCLN (σ).

Finalmente, la prueba de conjunto superior (Tabla 57) muestra que el modelo LSTM GJR-GARCH VIX ocupa la primera posición, seguido del LSTM EGARCH, confirmando que la integración entre memoria recurrente y estructuras asimétricas ofrece la mejor capacidad predictiva fuera de muestra. Las diferencias estadísticas entre los modelos son leves, pero suficientes para concluir que las configuraciones híbridas presentan un mejor balance entre precisión, estabilidad y capacidad de generalización por sobre los modelos clásicos.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,11663	1
LSTM EGARCH	2	-0,49927	1
LSTM GJR-GARCH	3	0,123308	0,9992
LSTM GARCH VIX	4	0,867682	0,7192
LSTM GARCH	5	1,117508	0,5266

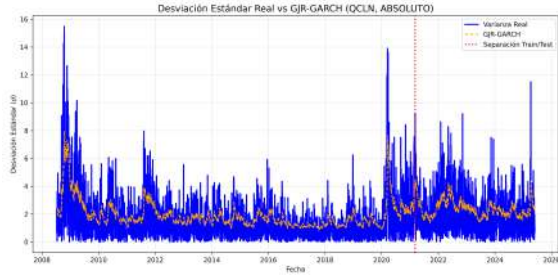
Cuadro 57: Conjunto superior de modelos para desviación estándar de QCLN.



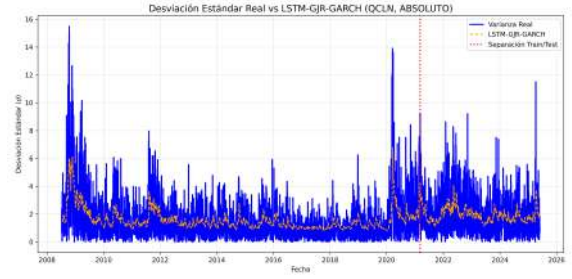
(a) Desviación estándar - GARCH



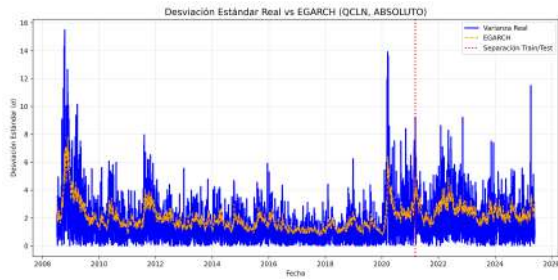
(b) Desviación estándar - LSTM GARCH



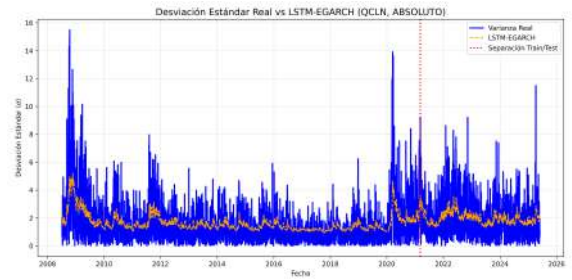
(c) Desviación estándar - GJR-GARCH



(d) Desviación estándar - LSTM GJR-GARCH



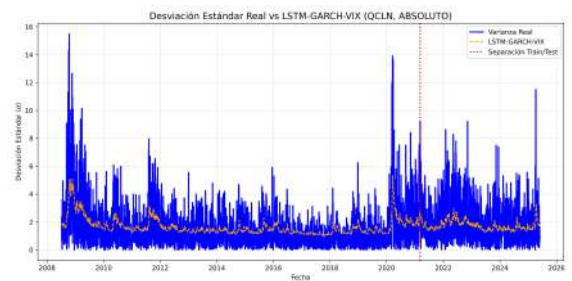
(e) Desviación estándar - EGARCH



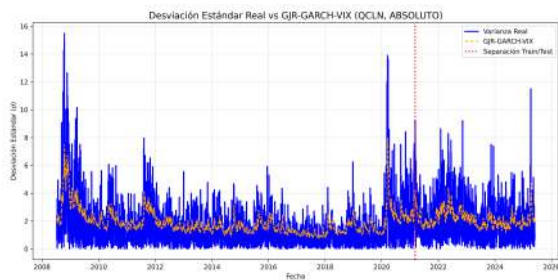
(f) Desviación estándar - LSTM EGARCH



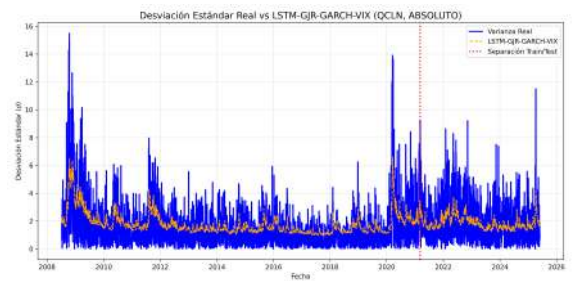
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.7. TAN

5.7.1. Entrenamiento modelos familia GARCH TAN (σ^2)

Los resultados de la optimización, presentados en la Tabla 58, indican que los modelos que incorporan la variable VIX logran los mejores valores de AIC, siendo el GJR-GARCH VIX el más eficiente con una especificación de orden (5,5,1,5). Esto sugiere que la inclusión de la volatilidad implícita del mercado aporta información relevante para explicar la dinámica condicional de la varianza en el ETF TAN.

Modelo	p	o	q	vix	AIC
GARCH	1	-	1	-	14671,1569
GJR-GARCH	1	1	1	-	14669,7038
EGARCH	3	4	4	-	14685,4341
GARCH VIX	1	-	1	5	14663,4821
GJR-GARCH VIX	5	5	1	5	14660,9174

Cuadro 58: Especificación de rezagos óptimos para modelos de familia GARCH TAN.

En la Tabla 59 se observa que, en términos de métricas de desempeño, las diferencias entre los modelos son moderadas. El GJR-GARCH presenta el mejor MSE y un rendimiento ligeramente superior en comparación con las demás especificaciones. Asimismo, el modelo EGARCH exhibe el mayor valor de MDA, lo que indica una mejor capacidad para capturar la dirección de los movimientos de volatilidad.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	19,954	516,139	9,049	1069694,342	108,080	2,765	0,296
GJR-GARCH	20,291	513,968	9,042	954060,438	105,949	2,764	0,303
EGARCH	19,617	518,314	8,934	1028885,690	106,891	2,760	0,356
GARCH VIX	19,960	516,098	9,054	1076929,083	108,339	2,766	0,295
GJR-GARCH VIX	19,050	521,971	9,054	1043097,162	106,972	2,757	0,339

Cuadro 59: Métricas de desempeño en entrenamiento familia GARCH TAN (σ^2).

5.7.2. Entrenamiento modelos híbridos Varianza TAN

Los hiperparámetros óptimos de los modelos híbridos se resumen en la Tabla 60. En general, las configuraciones muestran arquitecturas equilibradas, con entre 2 y 13 rezagos, tasas de apren-

dizaje reducidas (0,0001-0,001) y valores moderados de descarte aleatorio. Estas configuraciones apuntan a un entrenamiento estable, con una capacidad suficiente para capturar la persistencia temporal de la volatilidad.

Modelo LSTM	Rezagos	Neuro- nas C1	Neuro- nas C2	Neuro- nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	13	24	24	20	0	0,001	64	70
GJR-GARCH	3	12	12	2	0,2	0,01	64	30
EGARCH	2	20	2	12	0,2	0,01	64	60
GARCH VIX	9	4	50	20	0	0,01	64	60
GJR-GARCH VIX	2	12	2	24	0,2	0,01	64	50

Cuadro 60: Hiperparámetros óptimos para varianza de TAN.

Como se observa en la Tabla 61, los modelos híbridos presentan una mejora leve pero consistente en comparación con los modelos tradicionales. El LSTM GJR-GARCH VIX se posiciona como el mejor modelo dentro de este grupo, con el menor MSE (506,07), el menor HMAE (99,95) y un valor de R^2 de 21,52 %. Este desempeño sugiere que la combinación entre el componente asimétrico y la red LSTM permite una mayor sensibilidad ante choques negativos y una mejor adaptación a los cambios estructurales en la varianza condicional.

En contraste, los modelos LSTM EGARCH y LSTM GARCH VIX presentan resultados similares, aunque con una ligera pérdida de precisión en las métricas asociadas a errores absolutos. Esto indica que, si bien las redes neuronales aportan mejoras en la capacidad predictiva, la estructura base del modelo GJR-GARCH continúa siendo la más efectiva al capturar los patrones de volatilidad.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	18,141	527,832	8,609	665467,934	100,674	2,797	0,453
LSTM GJR-GARCH	20,739	511,079	8,809	875489,115	105,745	2,789	0,445
LSTM EGARCH	18,282	526,922	8,568	725782,333	102,393	2,817	0,467
LSTM GARCH VIX	18,437	525,922	8,572	742132,217	103,426	2,789	0,448
LSTM GJR-GARCH VIX	21,515	506,073	8,583	601747,788	99,950	2,798	0,467

Cuadro 61: Métricas de desempeño en entrenamiento modelos híbridos QCLN (σ).

5.7.3. Comparación de modelos fuera de muestra para Varianza TAN

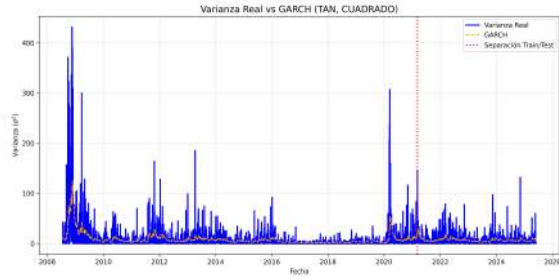
La Tabla 62 muestra que los modelos híbridos logran superar a los modelos tradicionales en la mayoría de las métricas fuera de muestra. El modelo LSTM GJR-GARCH VIX obtiene el mejor desempeño global, con un R^2 de 5,04 %, un MSE de 139,04 y el menor HMAE (275,85), confirmando su capacidad para generalizar la estructura de la volatilidad observada en el entrenamiento. Además, presenta una de las mayores tasas de acierto direccional (MDA = 0,468), lo que refuerza su eficacia para anticipar los movimientos de la varianza.

En comparación, el modelo LSTM GJR-GARCH logra resultados muy cercanos, mientras que el LSTM GARCH VIX ofrece una mejora adicional en estabilidad (menor HMSE). Los modelos clásicos, aunque estables, muestran menor capacidad predictiva y menor sensibilidad ante las fluctuaciones del mercado, lo que refuerza la superioridad del enfoque híbrido.

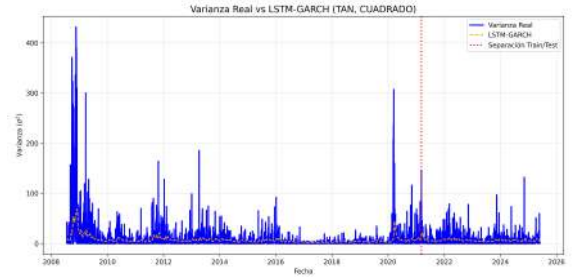
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	1,397	144,372	7,101	8817042,887	318,971	2,887	0,302
GJR-GARCH	2,076	143,377	7,053	8394258,405	312,896	2,883	0,301
EGARCH	2,549	142,684	7,220	9374641,133	336,552	2,882	0,353
GARCH VIX	1,346	144,446	7,111	8855099,779	319,841	2,887	0,301
GJR-GARCH VIX	3,054	141,945	7,058	8608655,376	320,710	2,884	0,372
LSTM GARCH	2,824	142,282	6,870	7440720,706	299,182	2,876	0,461
LSTM GJR-GARCH	3,745	140,934	6,806	7524664,109	303,295	2,877	0,468
LSTM EGARCH	3,226	141,694	6,729	6843927,368	296,727	2,881	0,457
LSTM GARCH VIX	4,517	139,803	6,841	6265373,162	281,741	2,880	0,443
LSTM GJR-GARCH VIX	5,036	139,044	6,574	6000262,163	275,846	2,866	0,468

Cuadro 62: Resultados de métricas de desempeño TAN (σ^2).

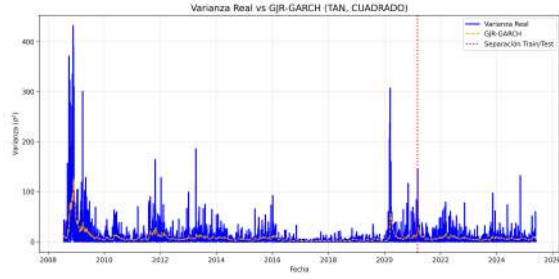
La prueba del conjunto superior de modelos (MCS), presentada en la Tabla 63, revela que, bajo un nivel de significancia del 10 %, los modelos GARCH y GARCH VIX son excluidos del conjunto, mientras que al aplicar un criterio más estricto del 5 %, ambos permanecen dentro del mismo. Este resultado evidencia que, aunque las arquitecturas híbridas más complejas alcanzan un mejor desempeño fuera de la muestra de entrenamiento, las mejoras obtenidas no son lo suficientemente consistentes ni estadísticamente sólidas como para justificar el descarte definitivo de los modelos econométricos clásicos.



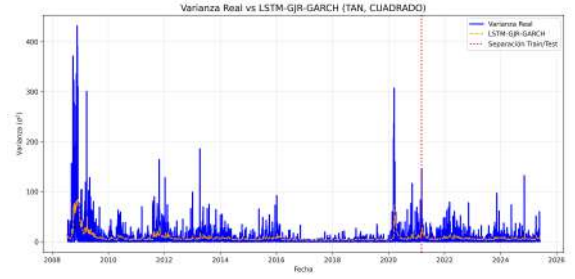
(a) Varianza - GARCH



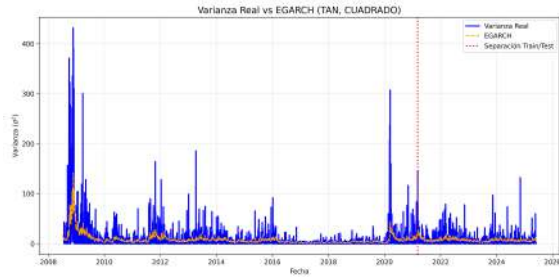
(b) Varianza - LSTM GARCH



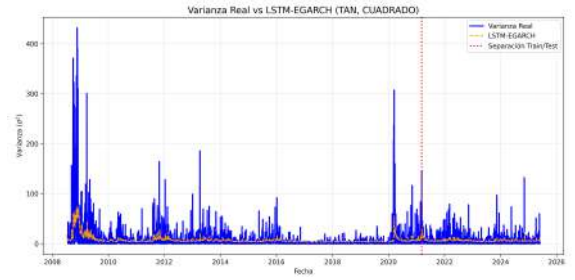
(c) Varianza - GJR-GARCH



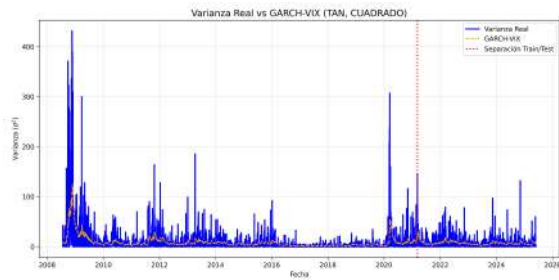
(d) Varianza - LSTM GJR-GARCH



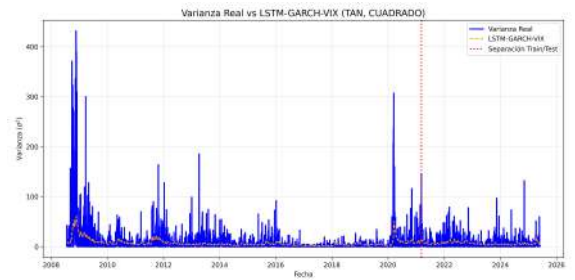
(e) Varianza - EGARCH



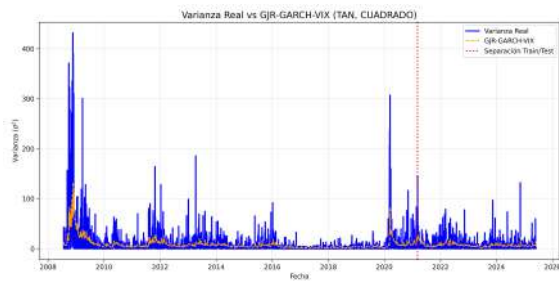
(f) Varianza - LSTM EGARCH



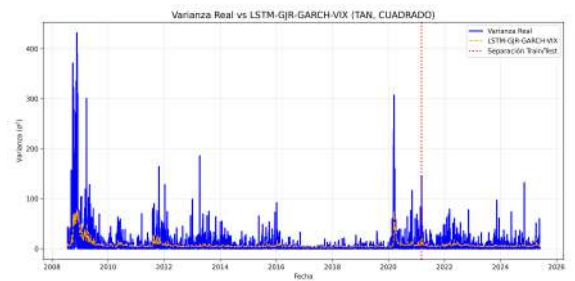
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 13: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF TAN (σ^2).

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-2,27369	1
LSTM GJR-GARCH	2	-1,90749	1
LSTM GARCH VIX	3	-1,36537	1
LSTM EGARCH	4	-0,32503	1
GJR-GARCH VIX	5	-0,10301	1
LSTM GARCH	6	0,221594	1
EGARCH	7	0,534982	0,9914
GJR-GARCH	8	1,584126	0,3844
GARCH	9	2,183284	0,0998
GARCH VIX	10	2,23149	0,0894

Cuadro 63: Conjunto superior de modelos para varianza de TAN.

5.7.4. Entrenamiento modelos familia GARCH TAN (σ)

Durante el entrenamiento (Tabla 64), los modelos clásicos de la familia GARCH presentan niveles similares de desempeño, con un coeficiente de determinación (R^2) en torno al 12 % y valores de MSE próximos a 3,9. Entre ellos, el modelo EGARCH destaca levemente con el mejor ajuste ($R^2 = 13,57\%$ y $MSE = 3,904$), lo que sugiere una ventaja marginal derivada de su capacidad para capturar asimetrías en la volatilidad. La inclusión del índice VIX no mejora el rendimiento, lo cual indica que la varianza condicional histórica explica la mayor parte de la volatilidad diaria observada.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	12,233	3,965	1,483	101,548	3,285	1,646	0,296
GJR-GARCH	12,542	3,951	1,480	99,439	3,274	1,645	0,303
EGARCH	13,569	3,904	1,475	100,385	3,273	1,645	0,355
GARCH VIX	12,163	3,968	1,483	101,802	3,288	1,646	0,295
GJR-GARCH VIX	12,178	3,967	1,480	100,470	3,271	1,644	0,339

Cuadro 64: Métricas de desempeño en entrenamiento familia GARCH TAN (σ).

5.7.5. Entrenamiento modelos híbridos Desviación estándar TAN

Los hiperparámetros óptimos del ETF TAN (Tabla 60) muestran configuraciones equilibradas entre complejidad y estabilidad. Los modelos sin VIX utilizan más rezagos (hasta 12), mientras que los que incorporan el índice reducen este número, lo que sugiere que el VIX aporta información adicional sobre la dinámica de la volatilidad.

En general, las arquitecturas presentan entre 8 y 24 neuronas por capa y un dropout de hasta 0.4, con tasas de aprendizaje de 0,01 o 0,001 según el modelo. Entre 40 y 70 épocas permiten un entrenamiento estable y con buena capacidad de generalización.

En la Tabla 66, correspondiente a los modelos híbridos, se observa una mejora generalizada en todas las métricas respecto de los modelos clásicos. En particular, los modelos LSTM GARCH VIX y LSTM GJR-GARCH VIX alcanzan los mejores desempeños en entrenamiento, con R^2 de 22,53 % y 23,71 %, respectivamente, y los menores errores cuadráticos medios ($MSE = 3,499$ y $3,446$). Estas mejoras reflejan la capacidad de las redes LSTM para modelar dependencias tempo-

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	12	24	16	16	0,4	0,01	128	50
GJR-GARCH	11	8	24	2	0,4	0,001	128	70
EGARCH	10	2	2	24	0,4	0,01	64	60
GARCH VIX	1	12	4	30	0,4	0,01	128	40
GJR-GARCH VIX	1	4	24	40	0	0,01	64	60

Cuadro 65: Hiperparámetros óptimos para desviación estándar de TAN.

rales no lineales y capturar la persistencia dinámica de la volatilidad. Adicionalmente, el indicador MDA, que mide la precisión en la dirección del cambio, se mantiene sobre 0,40, evidenciando una mejora en la capacidad direccional del modelo.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	18,466	3,683	1,298	58,887	2,593	1,629	0,423
LSTM GJR-GARCH	19,004	3,659	1,302	60,310	2,622	1,630	0,438
LSTM EGARCH	21,164	3,561	1,273	54,690	2,476	1,621	0,421
LSTM GARCH VIX	22,531	3,499	1,278	59,996	2,564	1,617	0,422
LSTM GJR-GARCH VIX	23,713	3,446	1,265	56,585	2,422	1,611	0,401

Cuadro 66: Métricas de desempeño en entrenamiento modelos híbridos TAN (σ).

5.7.6. Comparación de modelos fuera de muestra para Desviación Estándar TAN

En el análisis fuera de muestra (Tabla 67), las diferencias se vuelven más marcadas. Los modelos clásicos exhiben desempeños negativos en R^2 y errores sustancialmente mayores (MSE entre 3,09 y 3,18), mientras que los híbridos muestran una reducción significativa del error y una mejora en la capacidad explicativa. El modelo LSTM GARCH VIX presenta el menor MSE (2,594) y el mayor R^2 (4,67 %), seguido de cerca por el LSTM GJR-GARCH VIX (MSE = 2,606). Estas mejoras reflejan una capacidad más robusta para predecir la volatilidad durante períodos de alta variabilidad.

Al aplicar la prueba de Conjunto Superior de Modelos (Model Confidence Set, MCS) presentada en la Tabla 68, se observa que las diferencias entre modelos híbridos y clásicos son estadísticamente significativas al 5 %. Aunque los modelos híbridos con VIX se posicionan como los mejores en el ranking (posiciones 1 y 2), el valor- p superior a 0,05 indica que no existe evidencia suficiente

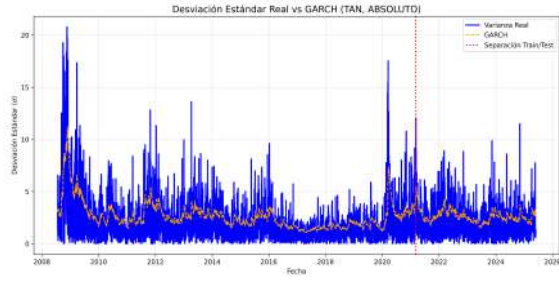
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-14,938 %	3,127	1,440	310,542	4,251	1,718	0,302
GJR-GARCH	-14,069 %	3,103	1,433	304,515	4,227	1,717	0,301
EGARCH	-16,954 %	3,182	1,466	327,848	4,386	1,723	0,353
GARCH VIX	-15,172 %	3,133	1,442	311,398	4,258	1,718	0,301
GJR-GARCH VIX	-13,642 %	3,092	1,436	312,278	4,253	1,717	0,372
LSTM GARCH	2,706 %	2,647	1,241	188,229	3,249	1,686	0,417
LSTM GJR-GARCH	2,533 %	2,652	1,245	190,263	3,266	1,686	0,431
LSTM EGARCH	2,747 %	2,646	1,238	169,666	3,144	1,686	0,416
LSTM GARCH VIX	4,665 %	2,594	1,227	173,959	3,177	1,681	0,428
LSTM GJR-GARCH VIX	4,217 %	2,606	1,223	156,649	3,051	1,686	0,435

Cuadro 67: Resultados de métricas de desempeño TAN (σ).

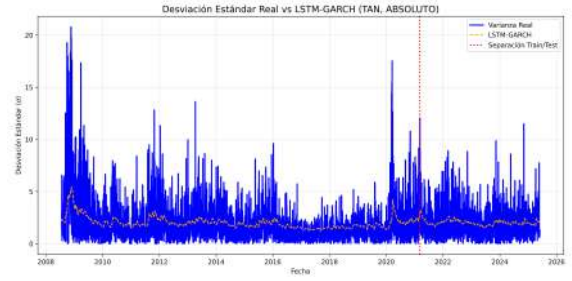
para rechazar la hipótesis de igualdad de desempeño entre los modelos de arquitectura LSTM.

Modelo	Posición	Estadístico	valor-p
LSTM GARCH VIX	1	-1,9001	1
LSTM GJR-GARCH VIX	2	-0,72696	1
LSTM EGARCH	3	0,991014	0,4802
LSTM GARCH	4	1,027739	0,4554
LSTM GJR-GARCH	5	1,174692	0,3674

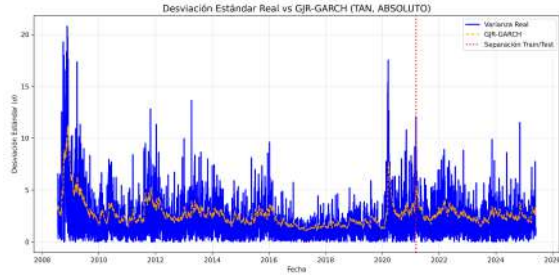
Cuadro 68: Conjunto superior de modelos para desviación estándar de TAN.



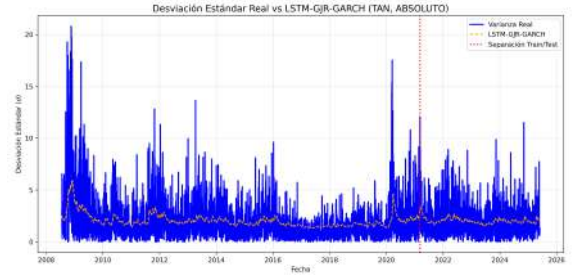
(a) Desviación estándar - GARCH



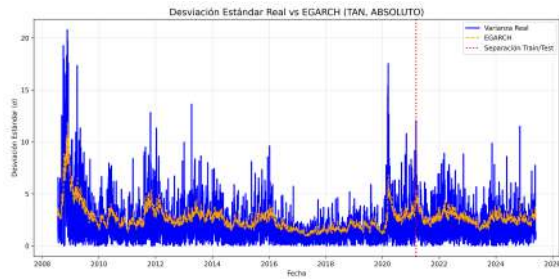
(b) Desviación estándar - LSTM GARCH



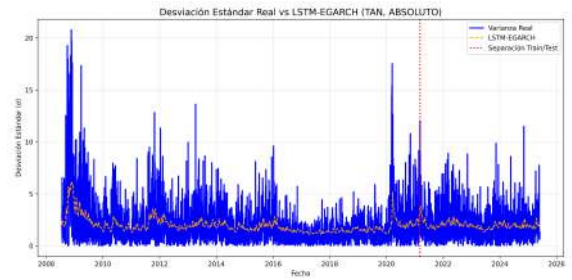
(c) Desviación estándar - GJR-GARCH



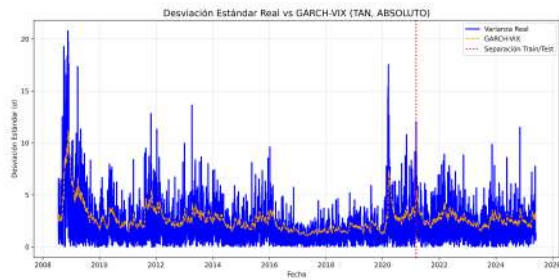
(d) Desviación estándar - LSTM GJR-GARCH



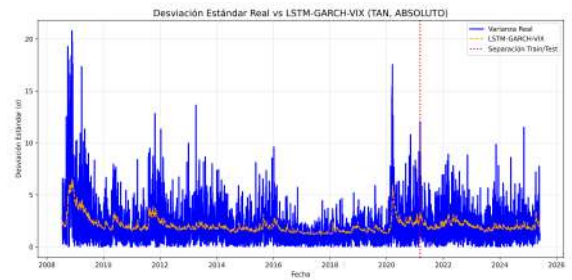
(e) Desviación estándar - EGARCH



(f) Desviación estándar - LSTM EGARCH



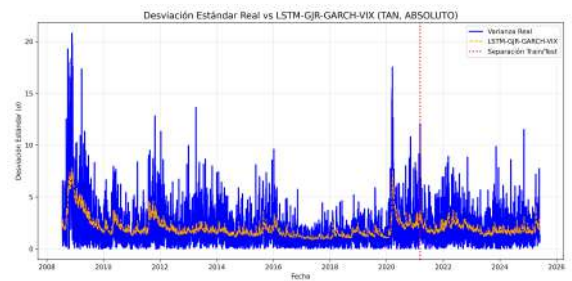
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.8. VDE

5.8.1. Entrenamiento modelos familia GARCH VDE (σ^2)

La Tabla 69 evidencia que el modelo GJR-GARCH VIX presenta el menor AIC, posicionándose como el más eficiente. Se observa además una tendencia hacia estructuras con múltiples rezagos tanto en los choques p como en la varianza pasada q , lo que refleja una alta persistencia en la volatilidad. Los modelos que incluyen la innovación de orden o presentan mejor rendimiento, por lo que parece existir el fenómeno del apalancamiento.

Modelo	p	o	q	vix	AIC
GARCH	2	-	2	-	11757,26735
GJR-GARCH	2	1	3	-	11697,71984
EGARCH	3	2	5	-	11689,95139
GARCH VIX	3	-	3	2	11762,54327
GJR-GARCH VIX	3	3	1	2	11698,21456

Cuadro 69: Especificación de rezagos óptimos para modelos de familia GARCH VDE.

Los resultados de la Tabla 70 muestran que los modelos de la familia GARCH aplicados al ETF VDE presentan un desempeño relativamente consistente, aunque con mejoras marginales entre las distintas variantes. Dentro del conjunto, el modelo EGARCH destaca por lograr el menor error cuadrático medio ($MSE = 222,229$) y el menor error absoluto medio ($MAE = 4,157$), alcanzando además el mayor coeficiente de determinación ($R^2 = 21,13\%$). Estos resultados sugieren que la estructura logarítmica del EGARCH es más efectiva para modelar la asimetría en la volatilidad del ETF, especialmente ante choques negativos en los retornos.

En cuanto a la incorporación del índice VIX como variable exógena, los modelos GARCH VIX y GJR-GARCH VIX no logran mejoras significativas respecto a sus contrapartes estándar, lo cual indica que la información implícita en la volatilidad del mercado no tiene un efecto determinante en la predicción de la varianza condicional de esta serie. No obstante, los valores de QLIKE cercanos entre los modelos reflejan una capacidad predictiva razonablemente estable en todo el conjunto.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	19,036	228,130	4,386	3748336,492	230,815	1,841	0,383
GJR-GARCH	20,888	222,910	4,324	3517065,625	228,246	1,824	0,404
EGARCH	21,130	222,229	4,157	4078124,441	234,555	1,816	0,426
GARCH VIX	21,090	222,342	4,269	3670743,569	228,512	1,836	0,379
GJR-GARCH VIX	19,417	227,055	4,350	3700424,604	230,478	1,819	0,416

Cuadro 70: Métricas de desempeño en entrenamiento familia GARCH VDE (σ^2).

5.8.2. Entrenamiento modelos híbridos Varianza VDE

La Tabla 71 presenta los hiperparámetros óptimos utilizados en los modelos híbridos LSTM-GARCH. En general, los modelos con una arquitectura más profunda (mayor número de neuronas por capa y múltiples rezagos) tienden a lograr mejores resultados, especialmente cuando se incluye una ligera regularización mediante *dropout* (0,2–0,4). Esto sugiere que la serie del ETF VDE contiene dependencias temporales no lineales que las redes neuronales logran capturar de forma más eficiente que los modelos econométricos tradicionales.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	14	24	30	8	0	0,01	64	40
GJR-GARCH	4	8	20	4	0,4	0,01	128	70
EGARCH	4	4	8	8	0	0,01	64	40
GARCH VIX	6	20	2	12	0,2	0,01	64	60
GJR-GARCH VIX	4	60	40	60	0,4	0,01	64	30

Cuadro 71: Hiperparámetros óptimos para varianza de VDE.

Los resultados de entrenamiento (Tabla 72) confirman estas mejoras. El modelo LSTM GARCH VIX alcanza el mejor desempeño general, con un R^2 de 26,25 % y el menor MSE (207,802), aunque con un MAE ligeramente superior debido a la mayor sensibilidad del modelo a los valores extremos. Los modelos LSTM EGARCH y LSTM GJR-GARCH VIX también obtienen resultados competitivos, manteniendo menores errores heterocedásticos (HMSE y HMAE), lo que evidencia una capacidad robusta de ajuste frente a la variabilidad de la serie.

Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	17,472	232,535	3,962	3392550,059	226,987	1,902	0,383
LSTM GJR-GARCH	17,344	232,896	4,110	4638805,044	269,863	1,922	0,483
LSTM EGARCH	21,348	221,614	4,029	2964532,920	210,439	1,833	0,475
LSTM GARCH VIX	26,250	207,802	4,491	6202475,049	310,262	1,895	0,461
LSTM GJR-GARCH VIX	22,707	217,784	4,204	4928534,229	269,307	1,849	0,445

Cuadro 72: Métricas de desempeño en entrenamiento modelos híbridos VDE (σ^2).

5.8.3. Comparación de modelos fuera de muestra para Varianza VDE

En el análisis fuera de muestra (Tabla 73), los modelos híbridos mantienen ventaja sobre los modelos clásicos. Los enfoques basados en LSTM reducen los errores de predicción y mejoran el ajuste explicativo de la varianza condicional. En particular, el modelo LSTM GJR-GARCH VIX obtiene el mejor resultado con un R^2 de 10,76 %, acompañado de los menores valores de MSE (37,567) y QLIKE (1,980), lo que confirma su mayor precisión y estabilidad predictiva.

Los resultados sugieren que la combinación de información exógena (VIX) y estructuras neuronales profundas permite capturar mejor las variaciones no lineales y dependientes del tiempo. En cambio, los modelos GARCH estándar mantienen un rendimiento limitado, con un R^2 cercano al 4,5 % y una precisión inferior en las métricas de error heterocedástico (HMSE y HMAE).

Modelo	$R^2 \%$	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	4,475	40,215	3,320	17948750,826	477,587	2,037	0,364
GJR-GARCH	4,661	40,137	3,296	13738200,768	431,684	2,032	0,405
EGARCH	6,332	39,433	3,238	13743011,354	420,821	2,023	0,423
GARCH VIX	4,917	40,029	3,296	16763785,301	462,082	2,017	0,367
GJR-GARCH VIX	3,752	40,519	3,298	13873193,776	425,199	2,031	0,416
LSTM GARCH	6,918	39,186	3,028	10864705,642	383,667	2,040	0,388
LSTM GJR-GARCH	6,818	39,229	3,139	12278276,108	423,440	2,036	0,468
LSTM EGARCH	7,343	39,007	3,083	11015354,327	389,527	2,031	0,462
LSTM GARCH VIX	10,224	37,795	3,397	19370402,845	503,524	1,992	0,450
LSTM GJR-GARCH VIX	10,764	37,567	3,280	17170007,815	463,858	1,980	0,421

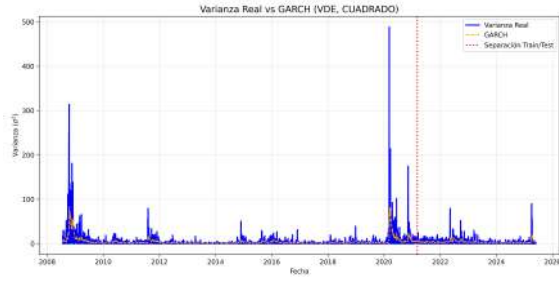
Cuadro 73: Resultados de métricas de desempeño VDE (σ^2).

La prueba de conjunto superior (Tabla 74) demuestra que la aparente superioridad de los modelos con arquitectura LSTM resulta ser un espejismo estadístico. Ninguno de los diez modelos evaluados alcanza un valor-p inferior al umbral de significancia del 5 %, lo que indica que las

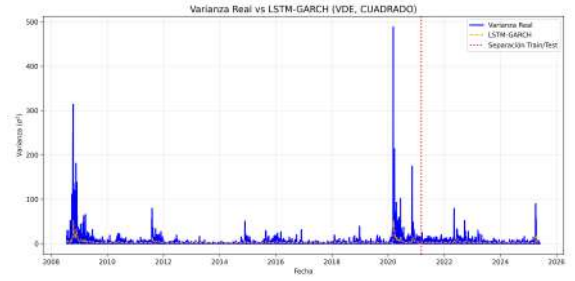
diferencias en desempeño no son estadísticamente significativas. En consecuencia, se infiere que, si bien los modelos híbridos pueden mostrar mejoras marginales en algunas métricas, estas no representan ventajas consistentes ni robustas frente a los modelos econométricos clásicos.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,81325	1
LSTM GARCH VIX	2	-1,68949	1
LSTM EGARCH	3	-0,57494	1
LSTM GARCH	4	-0,22052	1
LSTM GJR-GARCH	5	-0,15253	1
EGARCH	6	0,292503	0,9966
GJR-GARCH	7	1,222809	0,5194
GARCH VIX	8	1,483521	0,3402
GJR-GARCH VIX	9	1,574457	0,2922
GARCH	10	1,806291	0,2024

Cuadro 74: Conjunto superior de modelos para varianza de VDE.



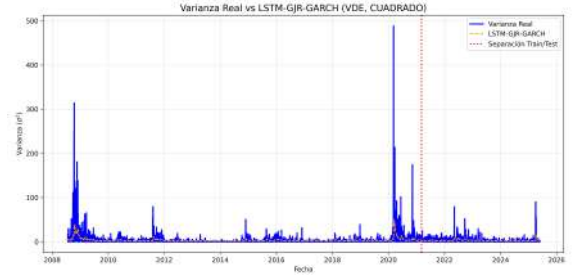
(a) Varianza - GARCH



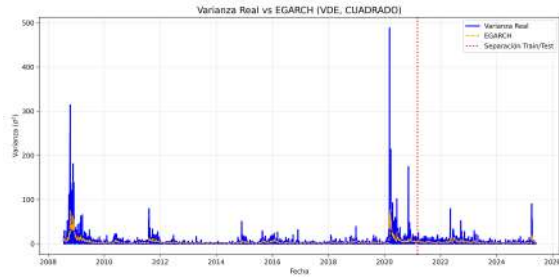
(b) Varianza - LSTM GARCH



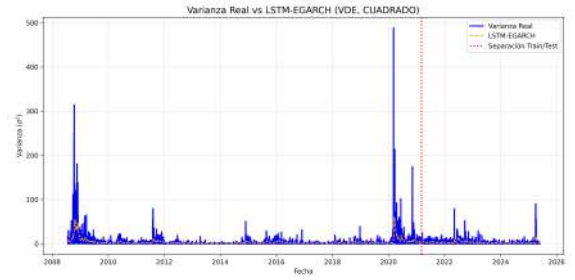
(c) Varianza - GJR-GARCH



(d) Varianza - LSTM GJR-GARCH



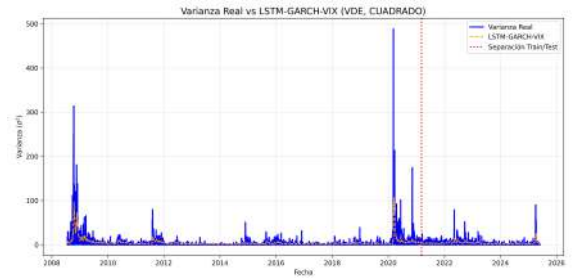
(e) Varianza - EGARCH



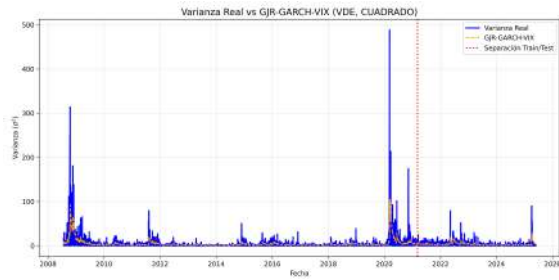
(f) Varianza - LSTM EGARCH



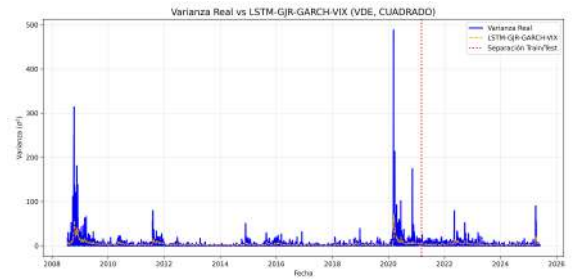
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 15: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF VDE (σ^2).

5.8.4. Entrenamiento modelos familia GARCH VDE (σ)

Los resultados de la Tabla 75 evidencian un desempeño consistente entre los modelos de la familia GARCH aplicados al ETF VDE, a pesar de solo aplicar raíz cuadrada sobre la serie, con mejoras graduales en la capacidad explicativa conforme aumenta la complejidad del modelo. El EGARCH presenta el mejor ajuste con un R^2 de 23,65 % y el menor error cuadrático medio (MSE = 1,876), lo que sugiere que la modelación logarítmica de la varianza permite capturar de forma más eficiente los efectos asimétricos de la volatilidad. La inclusión del índice VIX como variable exógena no muestra una mejora sustancial, indicando que la información de volatilidad implícita del mercado tiene un impacto limitado sobre la serie de este ETF.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	18,623	1,999	0,981	222,416	4,273	1,184	0,383
GJR-GARCH	20,435	1,955	0,968	219,942	4,225	1,178	0,404
EGARCH	23,650	1,876	0,954	226,262	4,220	1,177	0,426
GARCH VIX	21,441	1,930	0,970	220,165	4,246	1,182	0,379
GJR-GARCH VIX	19,591	1,975	0,968	222,180	4,222	1,178	0,416

Cuadro 75: Métricas de desempeño en entrenamiento familia GARCH VDE (σ).

5.8.5. Entrenamiento modelos Híbridos Desviación Estándar VDE

En la Tabla 76 se presentan los hiperparámetros óptimos utilizados para los modelos híbridos. Se observa una tendencia a arquitecturas más profundas, con múltiples capas y un nivel moderado de regularización, lo que refleja la necesidad de capturar relaciones no lineales más complejas en la volatilidad.

De acuerdo con la Tabla 77, los modelos híbridos superan claramente a los modelos tradicionales, logrando reducciones significativas en el MSE y el MAE. El modelo LSTM GJR-GARCH VIX alcanza el mejor desempeño general, con un R^2 de 30,51 % y el menor QLIKE (1,145), lo que evidencia su superioridad en la predicción de la desviación estándar. En general, los modelos híbridos con la variable exógena VIX presentan una mejor capacidad de generalización frente a los cambios de régimen en la volatilidad.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	9	16	12	16	0,4	0,01	64	50
GJR-GARCH	2	20	30	16	0	0,01	128	70
EGARCH	4	20	4	4	0,2	0,01	64	70
GARCH VIX	13	4	40	30	0	0,01	64	60
GJR-GARCH VIX	13	2	60	30	0	0,01	128	70

Cuadro 76: Hiperparámetros óptimos para desviación estándar de VDE.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	30,278	1,713	0,849	139,349	3,306	1,152	0,379
LSTM GJR-GARCH	27,799	1,774	0,868	148,055	3,489	1,161	0,417
LSTM EGARCH	27,897	1,771	0,847	126,138	3,186	1,152	0,457
LSTM GARCH VIX	29,254	1,738	0,855	130,492	3,284	1,152	0,416
LSTM GJR-GARCH VIX	30,508	1,707	0,846	123,694	3,184	1,145	0,457

Cuadro 77: Métricas de desempeño en entrenamiento modelos híbridos VDE (σ).

5.8.6. Comparación de modelos fuera de muestra para Desviación Estándar VDE

La Tabla 78 muestra que los modelos híbridos mantienen su superioridad fuera de muestra, con mejoras notorias en las métricas de error y estabilidad frente a los modelos GARCH convencionales. El modelo LSTM GJR-GARCH VIX alcanza los mejores resultados generales con un R^2 de 10,22 %, el menor MSE (1,252) y el menor QLIKE (1,231), demostrando una excelente capacidad de predicción y robustez frente a la variabilidad del mercado.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-8,660	1,515	0,984	467,441	5,112	1,270	0,364
GJR-GARCH	-6,906	1,491	0,972	421,889	4,939	1,264	0,405
EGARCH	-4,825	1,462	0,965	411,123	4,890	1,263	0,423
GARCH VIX	-7,120	1,494	0,979	452,049	5,056	1,267	0,367
GJR-GARCH VIX	-7,272	1,496	0,971	415,497	4,893	1,264	0,416
LSTM GARCH	8,182	1,280	0,840	250,584	3,747	1,238	0,368
LSTM GJR-GARCH	7,560	1,289	0,848	259,750	3,874	1,240	0,407
LSTM EGARCH	8,300	1,279	0,831	232,085	3,632	1,238	0,447
LSTM GARCH VIX	10,030	1,255	0,832	240,421	3,686	1,231	0,421
LSTM GJR-GARCH VIX	10,223	1,252	0,825	229,003	3,582	1,231	0,460

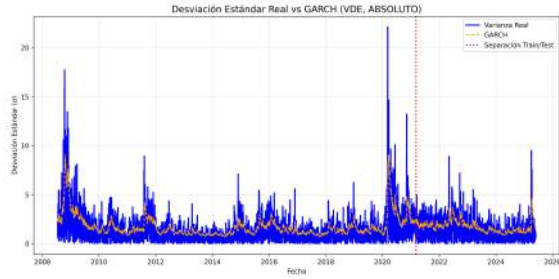
Cuadro 78: Resultados de métricas de desempeño VDE (σ).

La prueba de conjunto superior (Tabla 79) confirma la predominancia de los modelos híbri-

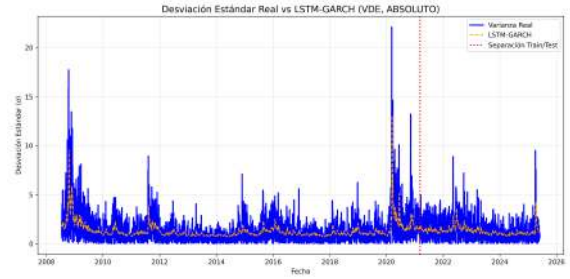
dos LSTM con componente exógeno —particularmente el LSTM GJR-GARCH VIX y el LSTM GARCH VIX—, los cuales lograron eliminar aproximadamente al 70 % de los modelos competidores. Este resultado respalda la relevancia del VIX como variable explicativa, evidenciando que la incorporación de información de volatilidad implícita mejora significativamente la capacidad predictiva de la desviación estándar.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,14954	1
LSTM GARCH VIX	2	-1,07218	1
LSTM EGARCH	3	1,626327	0,134

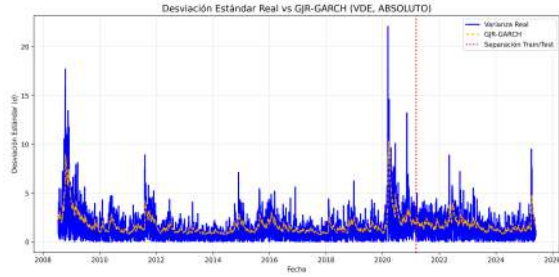
Cuadro 79: Conjunto superior de modelos para desviación estándar de VDE.



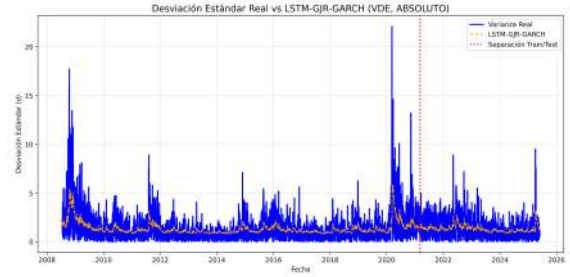
(a) Desviación estándar - GARCH



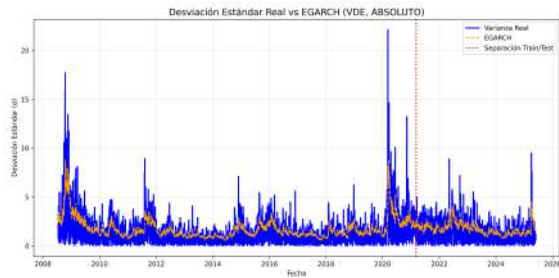
(b) Desviación estándar - LSTM GARCH



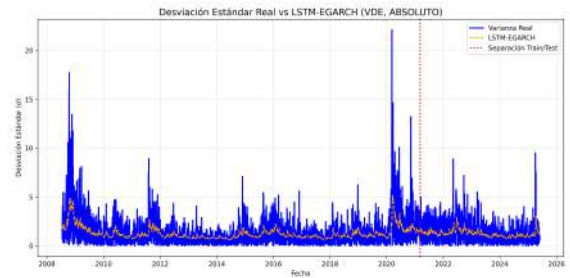
(c) Desviación estándar - GJR-GARCH



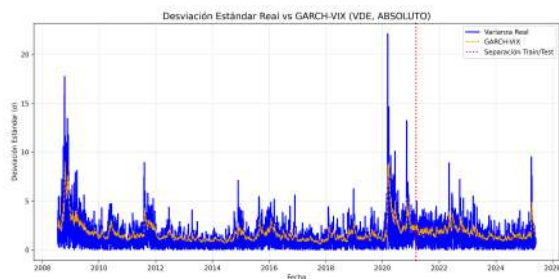
(d) Desviación estándar - LSTM GJR-GARCH



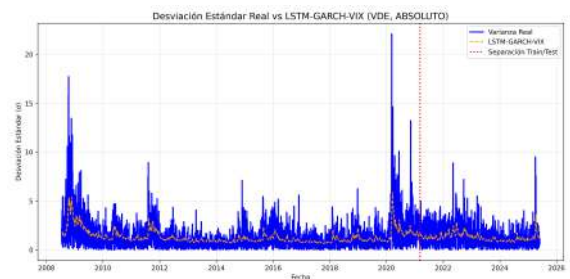
(e) Desviación estándar - EGARCH



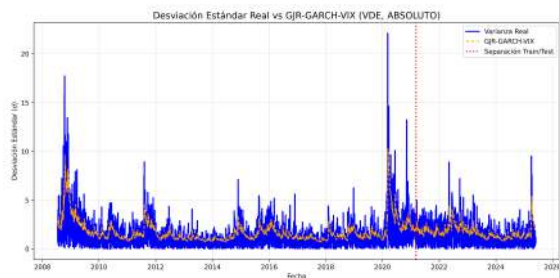
(f) Desviación estándar - LSTM EGARCH



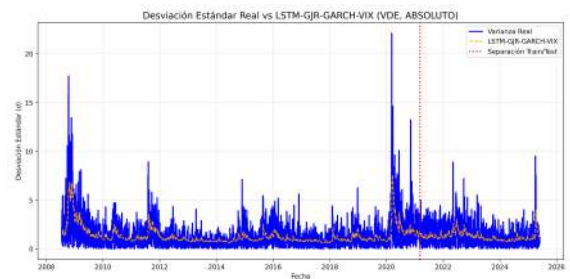
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.9. XLE

5.9.1. Entrenamiento modelos familia GARCH XLE (σ^2)

De acuerdo con la Tabla 80, los modelos GARCH de la familia XLE presentan un complejidad mixta ajustando p y q hasta orden 5 en el caso del EGARCH. Destacan el modelo EGARCH y el GJR-GARCH VIX por su menor AIC.

Modelo	p	o	q	vix	AIC
GARCH	2	-	3	-	11583,9038
GJR-GARCH	2	1	3	-	11520,9921
EGARCH	5	1	5	-	11514,7224
GARCH VIX	4	-	2	1	11587,3851
GJR-GARCH VIX	3	3	1	5	115179478

Cuadro 80: Especificación de rezagos óptimos para modelos de familia GARCH XLE.

La inclusión del VIX mejora levemente el ajuste en métricas como HMSE y HMAE, evidenciando una relación moderada entre la volatilidad implícita del mercado y la serie. Como se muestra en la Tabla 81, el modelo EGARCH alcanza el mejor desempeño dentro del grupo tradicional, con un R^2 de 21,26 % y el menor MSE, confirmando su capacidad para capturar asimetrías en la dinámica de la varianza.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	19,774	241,148	4,356	2314859,541	179,437	1,789	0,381
GJR-GARCH	20,939	237,648	4,299	1878869,658	172,319	1,771	0,405
EGARCH	21,259	236,686	4,112	1947663,956	174,272	1,764	0,424
GARCH VIX	19,999	240,472	4,246	2092615,123	178,136	1,785	0,359
GJR-GARCH VIX	20,062	240,283	4,319	1779413,601	171,558	1,766	0,408

Cuadro 81: Métricas de desempeño en entrenamiento familia GARCH XLE (σ^2).

5.9.2. Entrenamiento modelos Híbridos Varianza XLE

En la Tabla 82 se presentan los hiperparámetros óptimos de los modelos híbridos LSTM-GARCH. Se observa un predominio de configuraciones profundas con múltiples capas y tasas de aprendizaje altas, lo que favorece una convergencia estable durante el entrenamiento. Además, la mayoría de los modelos utiliza entre 40 y 50 épocas -con excepción del modelo GARCH con 70-, lo

que sugiere un equilibrio entre precisión y eficiencia computacional. Los modelos que incorporan la variable VIX tienden a requerir estructuras más complejas y mayor número de neuronas, lo que indica que la inclusión de información exógena incrementa la capacidad representativa, pero también la complejidad del ajuste.

Modelo LSTM	Rezagos	Neuro- nas C1	Neuro- nas C2	Neuro- nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	4	30	30	30	0,4	0,01	128	70
GJR-GARCH	4	16	12	4	0,4	0,01	128	40
EGARCH	3	24	24	20	0	0,001	64	50
GARCH VIX	3	12	16	30	0	0,001	128	40
GJR-GARCH VIX	5	40	8	40	0	0,01	64	50

Cuadro 82: Hiperparámetros óptimos para varianza de XLE.

La Tabla 83 presenta los resultados de entrenamiento de los modelos híbridos para la serie XLE. En términos generales, se observa un desempeño consistente entre las arquitecturas LSTM, con valores de R^2 cercanos al 20 %, lo que indica una baja capacidad explicativa de la volatilidad condicional.

El modelo LSTM GJR-GARCH VIX obtiene el mejor resultado global, alcanzando el menor MSE (229,795) y el valor de QLIKE más bajo (1,777), lo que refleja una mayor precisión en la estimación de la varianza. La inclusión del VIX parece contribuir positivamente a la estabilidad y ajuste del modelo, reduciendo el error cuadrático medio y mejorando la calidad de predicción.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	21,109	237,137	4,364	2412714,584	235,867	1,886	0,479
LSTM GJR-GARCH	17,097	249,195	4,042	1595937,882	196,662	1,913	0,471
LSTM EGARCH	21,165	236,968	3,833	888323,756	121,490	1,826	0,484
LSTM GARCH VIX	21,220	236,804	4,298	1997904,486	181,547	1,848	0,484
LSTM GJR-GARCH VIX	23,552	229,795	3,989	1281559,655	160,055	1,777	0,449

Cuadro 83: Métricas de desempeño en entrenamiento modelos híbridos XLE (σ^2).

5.9.3. Comparación de modelos fuera de muestra para Varianza XLE

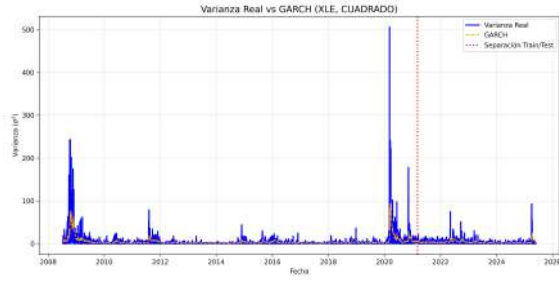
La Tabla 84 evidencia que los modelos híbridos superan consistentemente a los modelos clásicos en términos de bondad de ajuste y precisión predictiva. En particular, el modelo LSTM

GJR-GARCH VIX se destaca como el más eficiente, al presentar el menor error cuadrático medio y una alta estabilidad direccional reflejada en un MDA de 0,421. Asimismo, el modelo LSTM EGARCH sobresale por alcanzar los valores más bajos de HMSE y HMAE, indicando una mayor capacidad para capturar variaciones sutiles en la volatilidad diaria. Por su parte, el LSTM GARCH VIX obtiene el mejor desempeño en predicción de dirección, con un 50 % de aciertos, reforzando la utilidad de incorporar el VIX como variable explicativa en la modelación de la volatilidad.

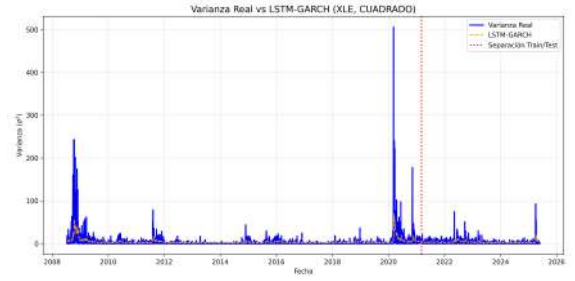
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	3,188	38,809	3,217	3635447,182	293,898	1,999	0,364
GJR-GARCH	4,205	38,401	3,183	4397070,469	308,916	2,000	0,402
EGARCH	6,408	37,518	3,137	5221531,708	318,301	1,987	0,412
GARCH VIX	5,800	37,762	3,165	3718955,423	292,749	1,980	0,347
GJR-GARCH VIX	3,245	38,786	3,189	4618706,166	309,497	1,998	0,403
LSTM GARCH	7,445	37,102	3,339	6525276,458	381,127	1,997	0,470
LSTM GJR-GARCH	6,712	37,396	3,003	4289856,538	309,661	2,002	0,465
LSTM EGARCH	6,424	37,511	2,829	2001939,489	208,529	2,166	0,496
LSTM GARCH VIX	8,560	36,655	3,204	4548415,720	307,754	1,985	0,500
LSTM GJR-GARCH VIX	10,127	36,027	2,907	3235275,838	265,991	1,940	0,421

Cuadro 84: Resultados de métricas de desempeño XLE (σ^2).

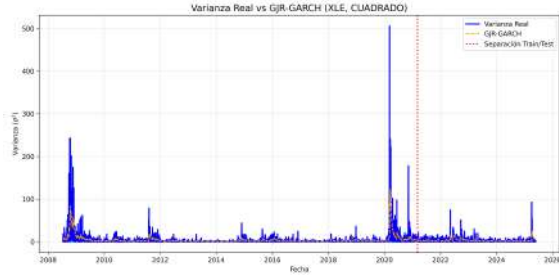
Finalmente, la prueba de conjunto superior (Tabla 85) demuestra que no existen diferencias estadísticas significativas entre los diez modelos evaluados. Esto implica que la mayor complejidad introducida por las variables exógenas y las arquitecturas LSTM no se traduce en una mejora estadísticamente verificable frente a los modelos clásicos al modelar la varianza. En otras palabras, aunque los modelos híbridos pueden mostrar ligeras ventajas numéricas, estas no son lo suficientemente robustas como para considerarlas superiores en términos de desempeño estadístico.



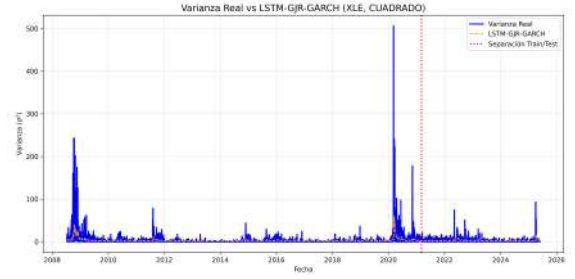
(a) Varianza - GARCH



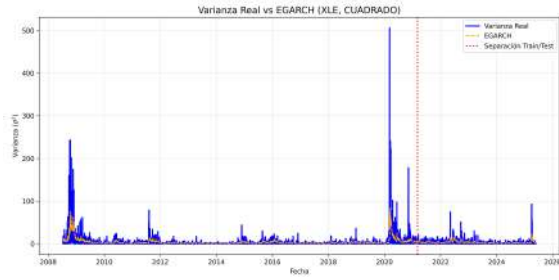
(b) Varianza - LSTM GARCH



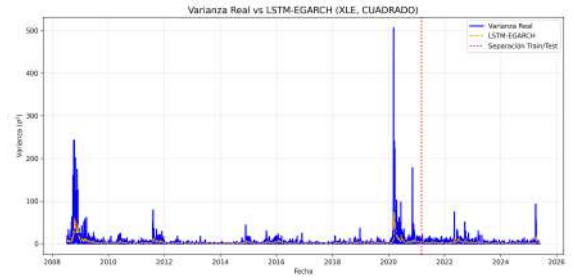
(c) Varianza - GJR-GARCH



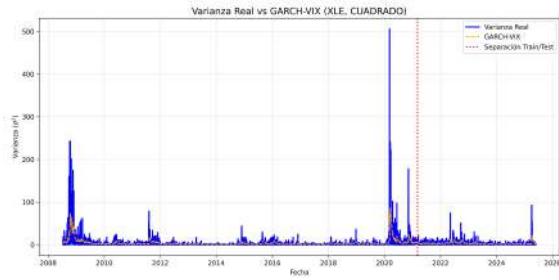
(d) Varianza - LSTM GJR-GARCH



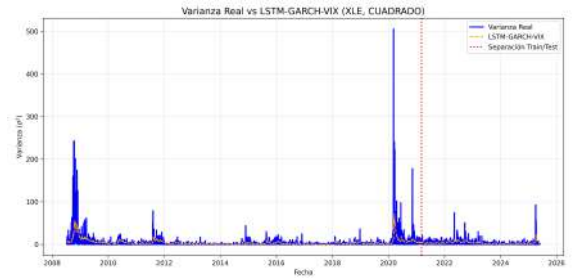
(e) Varianza - EGARCH



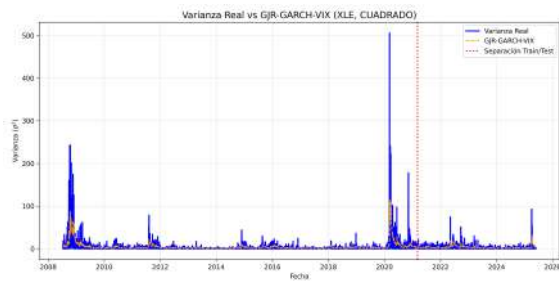
(f) Varianza - LSTM EGARCH



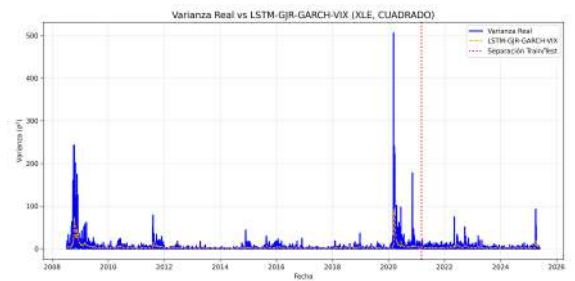
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 17: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XLE (σ^2).

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-2,64651	1
LSTM GARCH VIX	2	-1,46022	1
LSTM GARCH	3	-1,03675	1
LSTM GJR-GARCH	4	-0,38246	1
EGARCH	5	-0,17746	1
LSTM EGARCH	6	-0,13492	1
GARCH VIX	7	0,585152	0,9376
GJR-GARCH	8	1,159326	0,6056
GJR-GARCH VIX	9	1,425962	0,4116
GARCH	10	1,643823	0,2862

Cuadro 85: Conjunto superior de modelos para varianza de XLE.

5.9.4. Entrenamiento modelos familia GARCH XLE (σ)

En la Tabla 86 se aprecia que los modelos GARCH de la familia XLE presentan un desempeño estable, destacando el EGARCH con el mayor R^2 (24,99 %) y el menor MSE, lo que evidencia su capacidad para modelar la asimetría y los efectos de apalancamiento presentes en la serie. El resto de los modelos, incluyendo los que incorporan el VIX, mantienen un rendimiento similar, aunque con una ligera pérdida de ajuste en comparación con el EGARCH, lo que sugiere que la variable de volatilidad implícita no mejora sustancialmente el desempeño dentro del conjunto clásico.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	20,196	1,964	0,966	171,376	4,067	1,157	0,381
GJR-GARCH	21,656	1,928	0,951	164,380	4,007	1,152	0,405
EGARCH	24,993	1,846	0,937	166,333	4,007	1,150	0,424
GARCH VIX	22,332	1,912	0,954	170,106	4,052	1,155	0,359
GJR-GARCH VIX	21,016	1,944	0,952	163,633	4,000	1,151	0,408

Cuadro 86: Métricas de desempeño en entrenamiento familia GARCH XLE (σ).

5.9.5. Entrenamiento modelos Híbridos Desviación Estándar XLE

La Tabla 87 muestra los hiperparámetros óptimos obtenidos para los modelos híbridos. En general, las configuraciones seleccionadas presentan un número moderado de rezagos y tasas de aprendizaje reducidas, lo que favorece la estabilidad y la convergencia del proceso de entrenamiento. Además, se observa una preferencia de entre 40 y 60 épocas, lo que indica un equilibrio adecuado entre precisión y tiempo computacional. La tasa de descarte (*dropout*) se mantiene baja o nula en los modelos con componente exógeno, sugiriendo que la regularización es menos necesaria cuando la información del VIX contribuye a estabilizar el aprendizaje.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	9	16	20	8	0,4	0,01	64	40
GJR-GARCH	8	24	30	16	0,4	0,01	128	50
EGARCH	3	12	12	24	0,2	0,001	128	50
GARCH VIX	2	4	12	24	0,2	0,01	64	50
GJR-GARCH VIX	1	12	16	8	0	0,001	64	60

Cuadro 87: Hiperparámetros óptimos para desviación estándar de XLE.

Los resultados de desempeño en la Tabla 88 evidencian que los modelos híbridos superan ampliamente a los modelos GARCH tradicionales, siendo el LSTM GJR-GARCH el más destacado con un R^2 de 31,49% y el menor MSE (1,686). Esto indica que la red neuronal logra capturar de forma más eficiente las no linealidades y la memoria prolongada del proceso de volatilidad. Sin embargo, los modelos con VIX presentan valores cercanos en rendimiento, lo que sugiere que la inclusión del índice de volatilidad aporta estabilidad pero no una mejora sustancial del ajuste.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	30,896	1,701	0,829	93,833	3,062	1,124	0,388
LSTM GJR-GARCH	31,486	1,686	0,825	90,390	2,997	1,124	0,374
LSTM EGARCH	30,287	1,716	0,844	104,708	3,244	1,128	0,472
LSTM GARCH VIX	28,011	1,772	0,831	92,355	3,026	1,128	0,477
LSTM GJR-GARCH VIX	30,858	1,702	0,845	110,299	3,284	1,125	0,454

Cuadro 88: Métricas de desempeño en entrenamiento modelos híbridos XLE (σ).

5.9.6. Comparación de modelos fuera de muestra para Desviación Estándar XLE

La Tabla 89 presenta los resultados fuera de muestra, donde los modelos híbridos mantienen la superioridad mostrada en el entrenamiento. El LSTM GJR-GARCH VIX y el LSTM GARCH VIX destacan con los mejores valores de R^2 (10,10% y 9,90%, respectivamente) y los menores MSE y QLIKE, lo que evidencia su robustez frente a escenarios no vistos. La métrica MDA también muestra un rendimiento equilibrado, indicando estabilidad en la dirección de las predicciones.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-9,646	1,465	0,967	284,346	4,816	1,251	0,364
GJR-GARCH	-8,009	1,443	0,954	299,422	4,789	1,246	0,402
EGARCH	-5,842	1,414	0,950	308,761	4,811	1,246	0,412
GARCH VIX	-7,208	1,432	0,957	283,253	4,788	1,248	0,347
GJR-GARCH VIX	-8,339	1,447	0,954	300,040	4,771	1,246	0,403
LSTM GARCH	7,091	1,241	0,822	161,874	3,564	1,221	0,409
LSTM GJR-GARCH	6,818	1,245	0,819	159,325	3,494	1,223	0,401
LSTM EGARCH	7,962	1,230	0,830	178,264	3,736	1,219	0,504
LSTM GARCH VIX	9,904	1,204	0,804	144,778	3,352	1,213	0,477
LSTM GJR-GARCH VIX	10,105	1,201	0,822	170,368	3,646	1,210	0,423

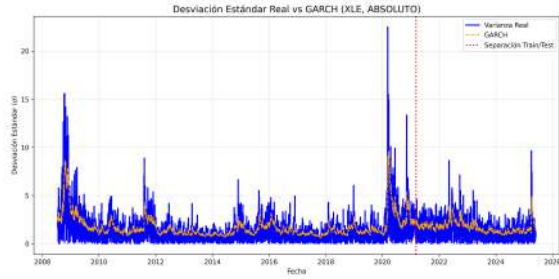
Cuadro 89: Resultados de métricas de desempeño XLE (σ).

El conjunto superior (Tabla 90) confirma estadísticamente estos hallazgos, situando al modelo

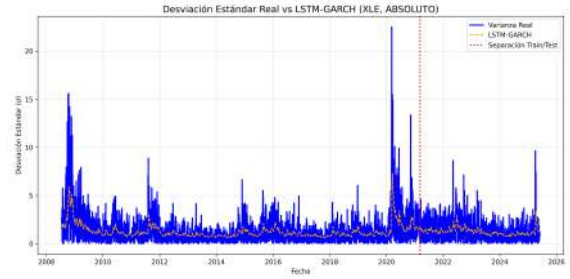
LSTM GJR-GARCH VIX como el mejor del grupo, seguido de cerca por el LSTM GARCH VIX. En conjunto, estos resultados reafirman la ventaja de las arquitecturas híbridas sobre los modelos puramente econométricos para capturar la desviación estándar del ETF XLE.

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-2,54183	1
LSTM GARCH VIX	2	-1,82215	1
LSTM EGARCH	3	0,677073	0,8226
LSTM GARCH	4	1,77325	0,1588
LSTM GJR-GARCH	5	2,156537	0,0708

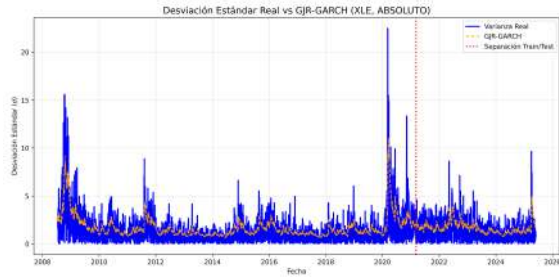
Cuadro 90: Conjunto superior de modelos para desviación estándar de XLE.



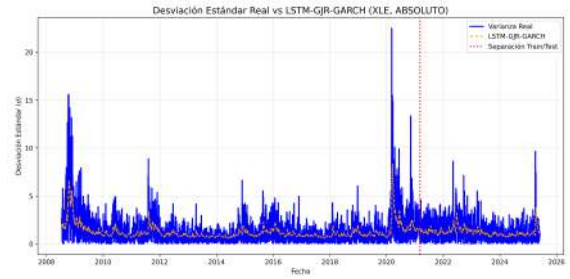
(a) Desviación estándar - GARCH



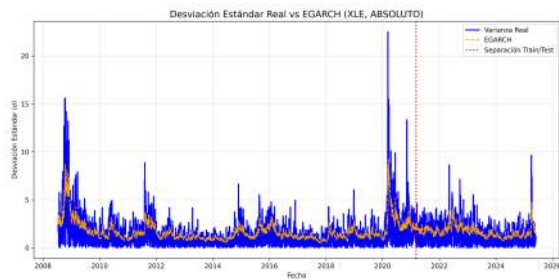
(b) Desviación estándar - LSTM GARCH



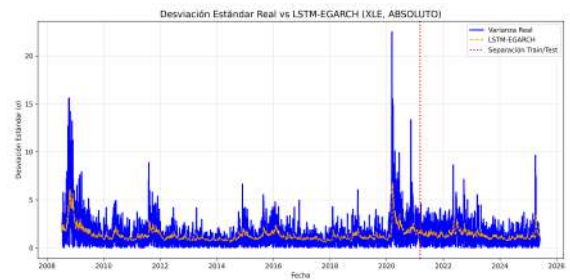
(c) Desviación estándar - GJR-GARCH



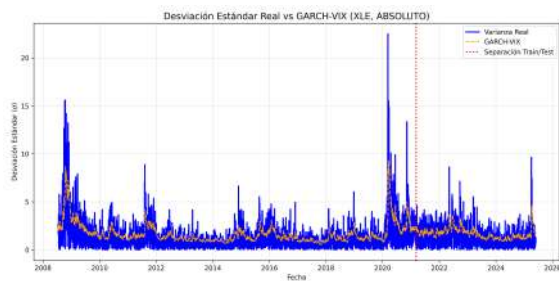
(d) Desviación estándar - LSTM GJR-GARCH



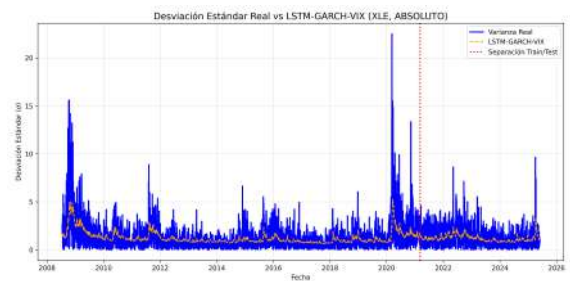
(e) Desviación estándar - EGARCH



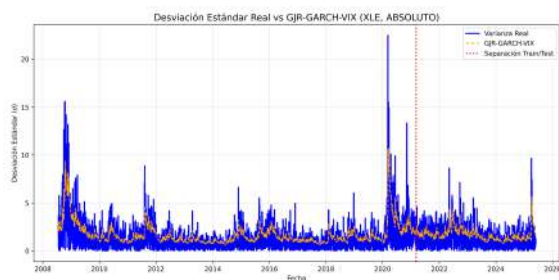
(f) Desviación estándar - LSTM EGARCH



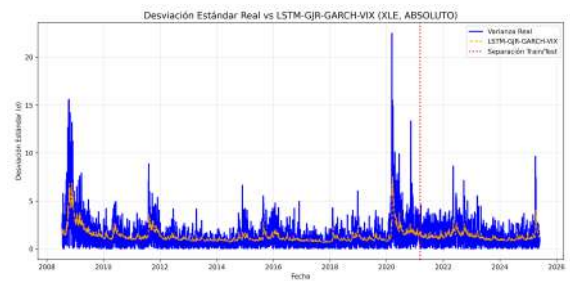
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

5.10. XOP

5.10.1. Entrenamiento modelos familia GARCH XOP (σ^2)

En la Tabla 92 se presentan los resultados obtenidos en el entrenamiento de la familia GARCH para el ETF XOP. Se observa que el modelo EGARCH logra el mejor desempeño, alcanzando el menor valor de AIC (14094,39) y un R^2 superior al resto de las especificaciones. Este resultado evidencia la existencia de asimetrías en la volatilidad, donde los choques negativos tienen un efecto más pronunciado que los positivos, fenómeno que es capturado eficientemente por la estructura logarítmica del EGARCH. Además, los modelos que incorporan la variable exógena VIX no muestran mejoras significativas en el ajuste, lo que sugiere que la volatilidad implícita del mercado ya está parcialmente internalizada en la dinámica de retornos del activo.

Modelo	p	o	q	vix	AIC
GARCH	3	-	3	-	14195,5074
GJR-GARCH	2	1	3	-	14126,0616
EGARCH	5	5	3	-	14094,3886
GARCH VIX	2	-	3	5	14085,7412
GJR-GARCH VIX	3	3	1	5	14080,2197

Cuadro 91: Especificación de rezagos óptimos para modelos de familia GARCH XOP.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	4,284	1767,017	8,266	1956447,752	134,951	2,608	0,377
GJR-GARCH	3,956	1773,080	8,282	1525143,182	127,913	2,589	0,395
EGARCH	9,493	1670,863	7,764	1642300,174	129,288	2,573	0,447
GARCH VIX	4,404	1764,810	8,241	1951789,676	135,482	2,609	0,302
GJR-GARCH VIX	4,772	1758,015	8,289	1586386,346	128,488	2,584	0,400

Cuadro 92: Métricas de desempeño en entrenamiento familia GARCH XOP (σ^2).

5.10.2. Entrenamiento modelos Híbridos Varianza XOP

Los hiperparámetros óptimos de los modelos híbridos LSTM-GARCH se muestran en la Tabla 93. Se observa que las configuraciones con mayores neuronas y un número moderado de rezagos obtuvieron los mejores resultados de entrenamiento. En la Tabla 94 se evidencia que los modelos híbridos superan en desempeño a los tradicionales, siendo el LSTM GJR-GARCH VIX

el más destacado con un R^2 de 8,81 % y un MSE de 1683,46. Este resultado confirma que la combinación de estructuras GARCH con redes LSTM mejora la capacidad para capturar patrones no lineales y dependencias de largo plazo en la volatilidad.

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	3	24	16	20	0,4	0,001	128	70
GJR-GARCH	2	4	20	16	0,4	0,01	64	40
EGARCH	2	12	12	4	0,2	0,01	128	30
GARCH VIX	1	16	24	40	0,2	0,001	64	50
GJR-GARCH VIX	5	16	16	60	0	0,01	128	70

Cuadro 93: Hiperparámetros óptimos para varianza de XOP.

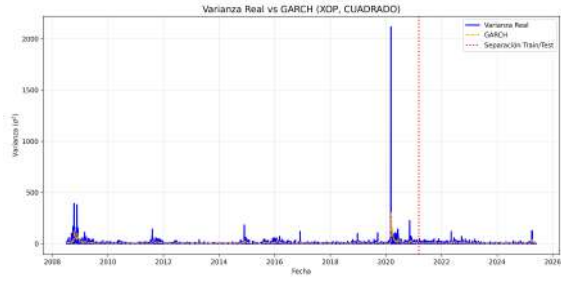
Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	4,409	1764,709	7,875	1208441,166	131,514	2,719	0,475
LSTM GJR-GARCH	7,303	1711,282	7,262	891131,113	102,597	2,617	0,455
LSTM EGARCH	5,671	1741,410	7,399	1097754,222	98,828	2,613	0,439
LSTM GARCH VIX	4,864	1756,320	7,859	1731239,937	123,250	2,637	0,506
LSTM GJR-GARCH VIX	8,810	1683,461	7,672	1769216,169	130,147	2,753	0,442

Cuadro 94: Métricas de desempeño en entrenamiento modelos híbridos XOP (σ^2).

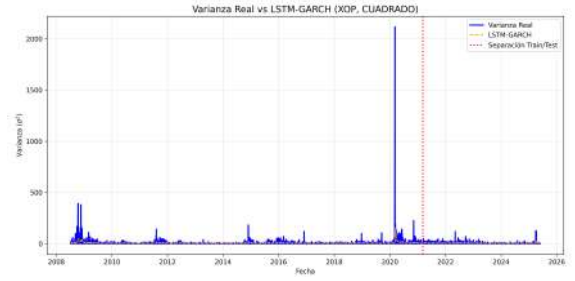
5.10.3. Comparación de modelos fuera de muestra para Varianza XOP

Los resultados fuera de muestra, mostrados en la Tabla 95, revelan que los modelos híbridos mantienen la superioridad obtenida en el entrenamiento. En particular, el LSTM GJR-GARCH VIX presenta el mejor desempeño con un R^2 de 10,23 %, acompañado de los menores valores de MSE y QLIKE. Este comportamiento sugiere que la red LSTM logra generalizar de manera eficiente la estructura de la volatilidad incluso en periodos no observados. La métrica MDA confirma la estabilidad direccional de las predicciones.

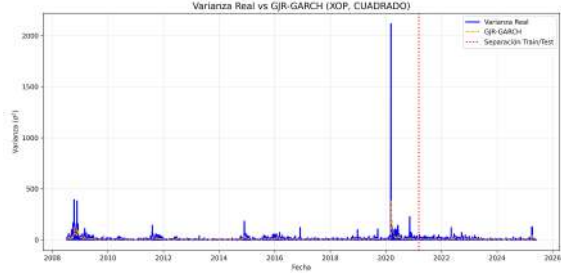
El conjunto superior de modelos (Tabla 96) valida estadísticamente estos resultados, situando al modelo LSTM GJR-GARCH como el mejor clasificado, seguido por su contraparte con VIX. En conjunto, los hallazgos demuestran que las arquitecturas híbridas superan claramente a los modelos econométricos tradicionales, consolidándose como una herramienta robusta para la estimación de la volatilidad del ETF XOP.



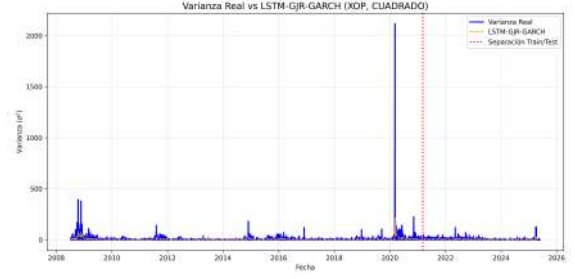
(a) Varianza - GARCH



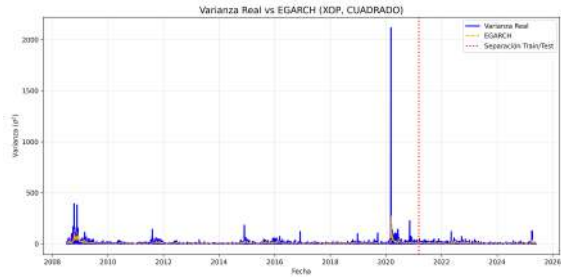
(b) Varianza - LSTM GARCH



(c) Varianza - GJR-GARCH



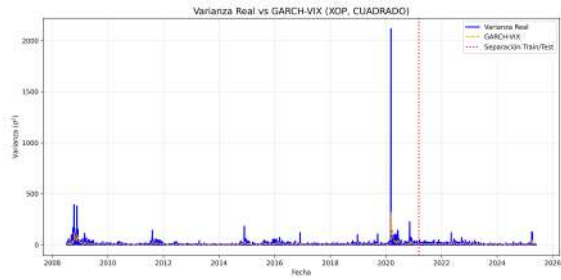
(d) Varianza - LSTM GJR-GARCH



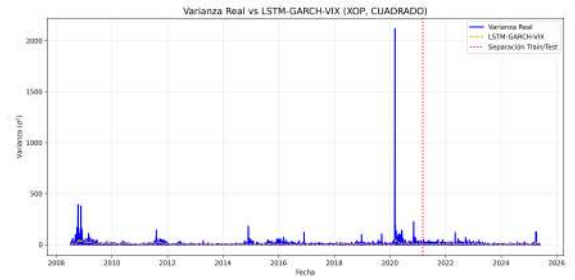
(e) Varianza - EGARCH



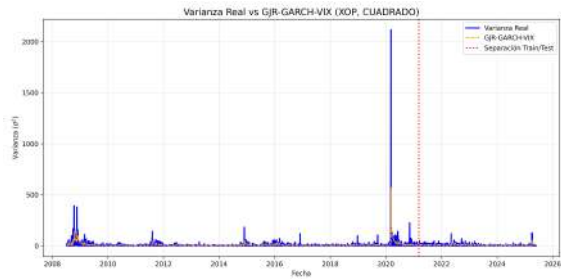
(f) Varianza - LSTM EGARCH



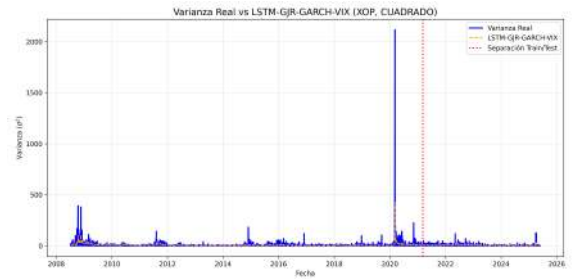
(g) Varianza - GARCH-VIX



(h) Varianza - LSTM GARCH-VIX



(i) Varianza - GJR-GARCH-VIX



(j) Varianza - LSTM GJR-GARCH-VIX

Figura 19: Comparación de modelos clásicos de volatilidad vs híbridos para el ETF XOP (σ^2).

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	3,861	110,353	5,492	105083936,935	889,084	2,515	0,376
GJR-GARCH	5,324	108,674	5,450	87570751,787	848,949	2,505	0,427
EGARCH	4,005	110,187	5,429	77557548,035	828,969	2,503	0,450
GARCH VIX	3,913	110,293	5,504	97389066,522	868,205	2,519	0,301
GJR-GARCH VIX	4,495	109,625	5,463	66799803,806	798,036	2,503	0,411
LSTM GARCH	6,780	107,002	5,553	74463714,505	941,406	2,550	0,516
LSTM GJR-GARCH	7,881	105,738	4,945	45856850,441	691,328	2,526	0,471
LSTM EGARCH	7,047	106,696	4,874	48422447,274	595,416	2,744	0,455
LSTM GARCH VIX	8,642	104,865	5,563	87036984,017	931,929	2,484	0,459
LSTM GJR-GARCH VIX	10,226	103,047	5,554	87229452,144	931,593	2,471	0,426

Cuadro 95: Resultados de métricas de desempeño XOP (σ^2).

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH	1	-2,2414	1
LSTM GJR-GARCH VIX	2	-1,59555	1
LSTM GARCH VIX	3	-1,05328	1
LSTM EGARCH	4	-0,50703	1
LSTM GARCH	5	-0,38813	1
LSTM GJR-GARCH	6	0,465337	0,981
GJR-GARCH VIX	7	0,997697	0,744
GARCH	8	1,6175	0,3176
EGARCH	9	1,671738	0,289
GARCH VIX	10	1,735592	0,2582

Cuadro 96: Conjunto superior de modelos para varianza de XOP.

5.10.4. Entrenamiento modelos familia GARCH XOP (σ)

En la Tabla 97 se presentan los resultados obtenidos en el entrenamiento de los modelos de la familia GARCH para el ETF XOP. El modelo EGARCH logra el mejor desempeño con un R^2 de 16,08 %, evidenciando una mayor capacidad para capturar las asimetrías en la volatilidad. Este comportamiento se explica por la naturaleza logarítmica del modelo, que permite representar de forma más precisa el impacto diferenciado de los choques negativos sobre la desviación estándar. En contraste, las versiones con el índice VIX no presentan mejoras significativas, lo que sugiere que el riesgo implícito de mercado ya se encuentra parcialmente reflejado en los retornos del activo.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	7,810	3,959	1,367	128,114	3,415	1,574	0,377
GJR-GARCH	7,950	3,953	1,354	121,206	3,353	1,568	0,394
EGARCH	16,086	3,603	1,325	122,615	3,335	1,564	0,447
GARCH VIX	7,455	3,974	1,367	128,628	3,423	1,574	0,302
GJR-GARCH VIX	8,526	3,928	1,357	121,786	3,349	1,567	0,399

Cuadro 97: Métricas de desempeño en entrenamiento familia GARCH XOP (σ).

5.10.5. Entrenamiento modelos Híbridos Desviación Estándar XOP

Los hiperparámetros óptimos utilizados para el entrenamiento de los modelos híbridos LSTM-GARCH se muestran en la Tabla 98. Se aprecia que las configuraciones con una estructura profunda (tres capas LSTM) y tasas de aprendizaje bajas ($\eta = 0,001$) obtienen los mejores resultados de ajuste. Según la Tabla 99, el desempeño de los modelos híbridos supera al de los modelos tradicionales, destacando el LSTM GARCH VIX con un R^2 de 21,96 % y el menor error cuadrático medio (MSE = 3,35). Esto confirma la capacidad de las redes LSTM para capturar relaciones no lineales y efectos de memoria extendida en la volatilidad.

5.10.6. Comparación de modelos fuera de muestra para Desviación Estándar XOP

Los resultados fuera de muestra, expuestos en la Tabla 100, muestran una mejora clara en el desempeño predictivo de los modelos híbridos. El modelo LSTM GJR-GARCH VIX destaca como el mejor, con un R^2 de 11,21 %, el menor error cuadrático medio (MSE = 2,02) y una alta precisión

Modelo LSTM	Rezagos	Neuro-nas C1	Neuro-nas C2	Neuro-nas C3	Descarte Aleatorio	Taza Aprendizaje	Lote	Épocas
GARCH	3	24	16	20	0,4	0,001	128	70
GJR-GARCH	3	24	16	20	0,4	0,001	128	70
EGARCH	3	24	30	20	0,2	0,001	64	50
GARCH VIX	3	8	24	40	0,4	0,01	128	50
GJR-GARCH VIX	2	8	30	24	0,2	0,001	128	70

Cuadro 98: Hiperparámetros óptimos para desviación estándar de XOP.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
LSTM GARCH	19,050	3,476	1,173	62,592	2,412147	1,546	0,455
LSTM GJR-GARCH	19,134	3,472978	1,178	64,294	2,489	1,547	0,444
LSTM EGARCH	20,680	3,407	1,177	68,657	2,533	1,541	0,457
LSTM GARCH VIX	21,961	3,351	1,176	70,751	2,534	1,541	0,374
LSTM GJR-GARCH VIX	20,419	3,417	1,1697	66,716	2,432	1,547	0,461

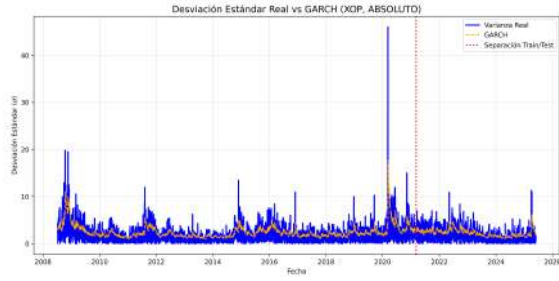
Cuadro 99: Métricas de desempeño en entrenamiento modelos híbridos XOP (σ).

direccional (MDA = 0,459). Estos resultados confirman la capacidad de las arquitecturas híbridas para capturar dinámicas no lineales y dependencias temporales en la volatilidad del activo.

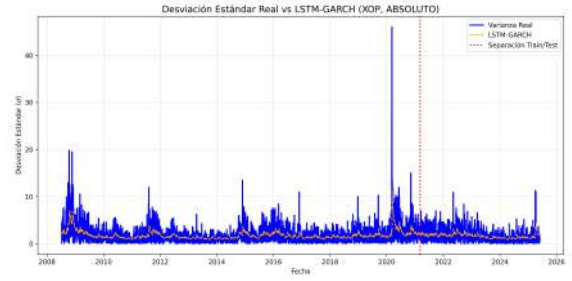
El conjunto superior de modelos (Tabla 101) valida estadísticamente estos hallazgos, ubicando al modelo LSTM GJR-GARCH VIX como el mejor clasificado, seguido por LSTM EGARCH y LSTM GARCH VIX. En conclusión, las redes híbridas LSTM-GARCH ofrecen un desempeño superior tanto en ajuste como en generalización, reafirmando su potencial para el modelado de la volatilidad en activos energéticos.

Modelo	R^2 %	MSE	MAE	HMSE	HMAE	QLIKE	MDA
GARCH	-10,147	2,508	1,244	876,679	6,238	1,519	0,376
GJR-GARCH	-7,647	2,451	1,228	836,731	6,145	1,513	0,427
EGARCH	-7,536	2,448	1,223	816,892	6,074	1,510	0,450
GARCH VIX	-10,293	2,511	1,246	855,856	6,210	1,520	0,301
GJR-GARCH VIX	-7,706	2,452	1,227	785,952	6,079	1,511	0,411
LSTM GARCH	8,341	2,087	1,042	442,817	4,496	1,485	0,478
LSTM GJR-GARCH	8,965	2,072	1,049	479,315	4,682	1,485	0,476
LSTM EGARCH	10,672	2,034	1,044	493,477	4,684	1,479	0,501
LSTM GARCH VIX	10,551	2,036	1,055	509,413	4,799	1,478	0,378
LSTM GJR-GARCH VIX	11,208	2,021	1,046	491,675	4,688	1,477	0,459

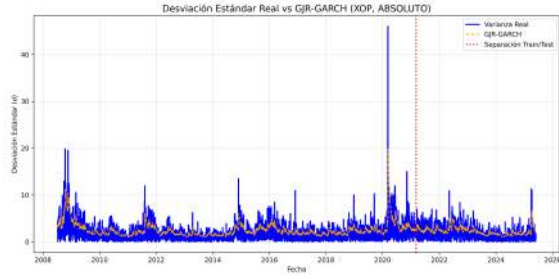
Cuadro 100: Resultados de métricas de desempeño XOP (σ).



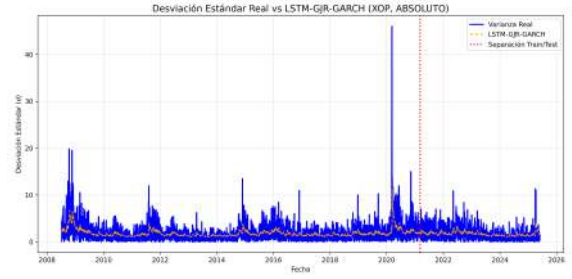
(a) Desviación estándar - GARCH



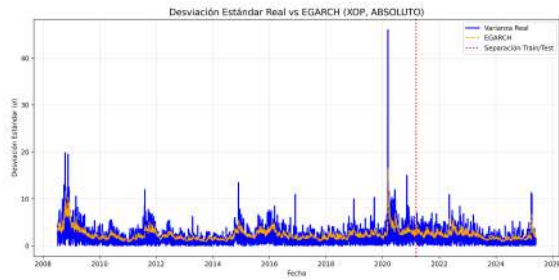
(b) Desviación estándar - LSTM GARCH



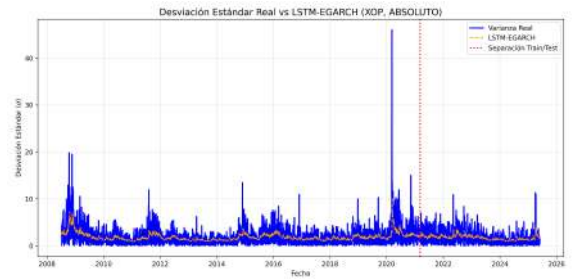
(c) Desviación estándar - GJR-GARCH



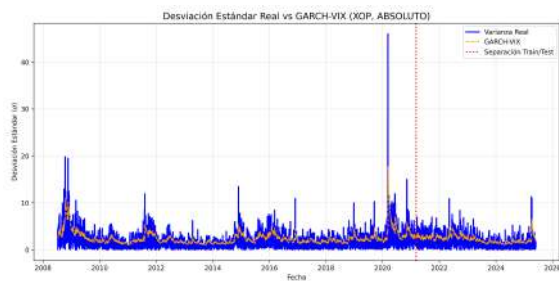
(d) Desviación estándar - LSTM GJR-GARCH



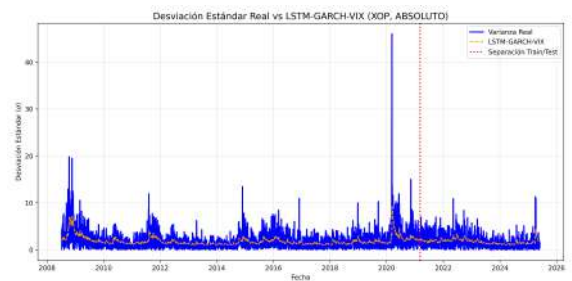
(e) Desviación estándar - EGARCH



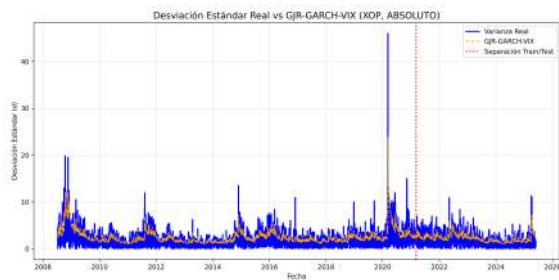
(f) Desviación estándar - LSTM EGARCH



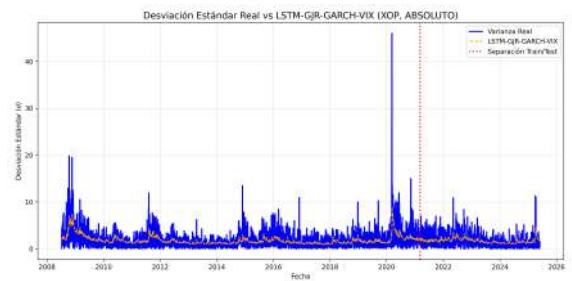
(g) Desviación estándar - GARCH-VIX



(h) Desviación estándar - LSTM GARCH-VIX



(i) Desviación estándar - GJR-GARCH-VIX



(j) Desviación estándar - LSTM GJR-GARCH-VIX

Modelo	Posición	Estadístico	valor-p
LSTM GJR-GARCH VIX	1	-1,64735	1
LSTM EGARCH	2	-0,97115	1
LSTM GARCH VIX	3	-0,86485	1
LSTM GARCH	4	2,111368	0,0568

Cuadro 101: Conjunto superior de modelos para desviación estándar de XOP.

6. Conclusiones

Los resultados empíricos confirman que los modelos híbridos LSTM–GARCH superaron de manera sistemática a las especificaciones econométricas tradicionales en la mayoría de las métricas de desempeño, con excepción del indicador QLIKE, donde las diferencias no resultaron estadísticamente significativas. Esta mejora fue especialmente consistente en la predicción de la desviación estándar, ámbito en el que las pruebas del *Model Confidence Set* (MCS) validaron la superioridad estadística de los modelos híbridos frente a los puramente econométricos.

Sin embargo, al considerar la varianza, la ventaja de los modelos híbridos se diluye. Esto se debe a que la transformación cuadrática amplifica las diferencias diarias, exacerbando la volatilidad observada durante periodos de alta incertidumbre, como la crisis subprime, la pandemia de COVID-19, la guerra en Ucrania o los episodios de tensiones comerciales entre Estados Unidos y China. En estos escenarios, la presencia de choques abruptos dificulta el proceso de aprendizaje y la posterior generalización, provocando desviaciones notables en las predicciones fuera de muestra.

En términos de desempeño por tipo de activo, los ETFs asociados a energías renovables (FAN, TAN, PBW, ICLN y QCLN) presentaron una volatilidad ligeramente superior y un comportamiento más heterogéneo, especialmente fuera del periodo de entrenamiento. Este resultado puede parecer contraintuitivo, dado que los ETFs del sector convencional podrían presumirse más volátiles debido a la influencia de conflictos geopolíticos como los de Ucrania y Medio Oriente. No obstante, los resultados indican que no existen diferencias significativas en la capacidad predictiva de los modelos entre ambos tipos de ETFs, lo que sugiere que la dinámica de volatilidad se comporta de manera similar en ambos sectores bajo las condiciones analizadas.

Respecto a la incorporación del índice VIX como variable exógena, su inclusión generó mejoras moderadas en la mayoría de los casos, especialmente en contextos de alta incertidumbre. Sin embargo, los resultados del MCS muestran que los modelos con VIX no lograron excluir a las versiones híbridas sin dicha variable en el caso de la desviación estándar, ni desplazar a los modelos econométricos clásicos en la predicción de la varianza. Esto evidencia que, aunque el VIX aporta información relevante sobre la volatilidad implícita del mercado, su efecto no es estructuralmente

dominante frente a los componentes históricos capturados por los propios modelos.

Finalmente, los hallazgos de este estudio sugieren que para la estimación de la desviación estándar, el mayor costo computacional y el tiempo de optimización asociados a los modelos híbridos se justifican plenamente debido a su capacidad para representar relaciones no lineales y mejorar la precisión predictiva. En cambio, para la predicción de la varianza, las ganancias de desempeño no compensan el incremento en complejidad, de modo que los modelos GARCH tradicionales continúan siendo una alternativa robusta, parsimoniosa y eficiente, particularmente en mercados energéticos consolidados y de menor volatilidad estructural.

Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, 12(7), e0180944. <https://doi.org/10.1371/journal.pone.0180944>
- BlackRock, Inc. (2024a). iShares Global Clean Energy ETF (ICLN) - Fund Overview.
- BlackRock, Inc. (2024b). iShares U.S. Energy ETF (IYE) - Fund Overview.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- Donaldson, R. G., & Kamstra, M. J. (1997). An artificial neural network–GARCH model for international stock return volatility. *Journal of Empirical Finance*, 4(1), 17-46. [https://doi.org/10.1016/S0927-5398\(96\)00011-4](https://doi.org/10.1016/S0927-5398(96)00011-4)
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1007. <https://doi.org/10.2307/1912773>
- Fabozzi, F. J. (2002). *Bond Markets, Analysis, and Strategies* (5th). Prentice Hall.
- First Trust Advisors. (2024a). First Trust Global Wind Energy ETF (FAN) - Fund Summary.
- First Trust Advisors. (2024b). First Trust NASDAQ Clean Edge Green Energy Index Fund (QCLN) - Fund Summary.
- Fuertes, A.-M., Izzeldin, M., & Kalotychou, E. (2009). On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting*, 25(2), 259-281. <https://doi.org/10.1016/j.ijforecast.2009.01.006>
- Gastineau, G. L. (2010). *The Exchange-Traded Funds Manual* (2.^a ed.). John Wiley Sons.

- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779-1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Graves, A. (2013). *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer. <https://doi.org/10.1007/978-3-642-24797-2>
- Grinblatt, M., & Titman, S. (2011). *Financial Markets and Corporate Strategy* (2nd). McGraw-Hill/Irwin.
- Gutiérrez, R. d. J. (2023). El uso de la volatilidad implícita en el modelado de la varianza condicional puede mejorar la predicción de la volatilidad y la estimación del VaR y el CVaR. *Economía: Teoría y Práctica*, 18(1), 173-199. <https://doi.org/10.24275/ETYPUAM/NE/582023/Jesus>
- Hansen, P. R., & Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873-889. <https://doi.org/10.1002/jae.800>
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453-497. <https://doi.org/10.3982/ECTA5771>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Invesco Ltd. (2024a). Invesco Solar ETF (TAN) - Product Details.
- Invesco Ltd. (2024b). Invesco WilderHill Clean Energy ETF (PBW) - Product Details.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259. [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5)
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), 159-178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>

- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- Mishkin, F. S. (2018). *The Economics of Money, Banking and Financial Markets* (12th). Pearson Education.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347-370. <https://doi.org/10.2307/2938260>
- Patton, A. J. (2011). Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics*, 160(1), 246-256. <https://doi.org/10.1016/j.jeconom.2010.03.034>
- Pesaran, M. H., & Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. *Journal of Business Economic Statistics*, 10(4), 461-465. <https://doi.org/10.2307/1391822>
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346. <https://doi.org/10.1093/biomet/75.2.335>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. En *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (pp. 318-362, Vol. 1). MIT Press.
- State Street Global Advisors. (2024a). Energy Select Sector SPDR Fund (XLE) - Fund Overview.
- State Street Global Advisors. (2024b). SPDR S&P Oil & Gas Exploration & Production ETF (XOP) - Fund Overview.
- Vanguard Group, Inc. (2024). Vanguard Energy ETF (VDE) - Fund Profile.
- Whaley, R. E. (2000). The investor fear gauge. *Journal of Portfolio Management*, 26(3), 12-17. <https://doi.org/10.3905/jpm.2000.319728>
- Zhang, W., Zhang, Y., Zhang, D., Guo, Y., & Wang, C. (2019). Out-of-sample prediction of the oil futures market: A new approach based on machine learning. *Energy Economics*, 81, 1025-1039. <https://doi.org/10.1016/j.eneco.2019.05.018>